Does Voluntary Information Disclosure Lead to Less Cooperation than Mandatory Disclosure? Evidence from a Sequential Prisoner's Dilemma Experiment

Georg Kirchsteiger, Tom Lenaerts, Rémi Suchon

Does Voluntary Information Disclosure Lead to Less Cooperation than Mandatory Disclosure? Evidence from a Sequential Prisoner's Dilemma Experiment

Georg Kirchsteiger
ECARES, Université libre de Bruxelles

Tom Lenaerts
Artificial Intelligence Laboratory, Vrije Universiteit Brussel
Machine Learning Group, Université libre de Bruxelles

Rémi Suchon
Anthropo Lab - Ethics EA 7446, Université Catholique de Lille

July 2022

ECARES working paper 2022-26

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Does voluntary information disclosure lead to less cooperation than mandatory disclosure ? Evidence from a sequential prisoner's dilemma experiment.*

Georg Kirchsteiger†, Tom Lenaerts‡, Rémi Suchon§

July 17, 2022

Abstract

In sequential social dilemmas with stranger matching, initiating cooperation is inherently risky for the first mover. The disclosure of the second mover's past actions may be necessary to instigate cooperation. We experimentally compare the effect of mandatory and voluntary disclosure with non disclosure in a sequential prisoner's dilemma situation. Our results confirm the positive effects of disclosure on cooperation. We also find that voluntary disclosure is as effective as mandatory one, which is surprising given the results of existing literature on this topic. With voluntary disclosure, second movers with a good track record decided to disclose because they expect that not disclosing signals non-cooperativeness. First movers interpret non-disclosure correctly as a signal of non-cooperativeness. Therefore, they cooperate less than half as often when the second mover does not disclose.

*We thank Elias Fernandez and Antoine Deplancke for helping with the experiment in Brussels and Lille. This project received financial support from Fonds de la Recherche Scientifique (FNRS, PDR T014318F).

†ECARES, Université Libre de Bruxelles, Bruxelles, Belgium - georg.kirchsteiger@ulb.be

‡Artificial Intelligence Laboratory, Vrije Universiteit Brussel & Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium - tom.lenaerts@ulb.be

§ANTHROPO LAB – ETHICS EA 7446, Université Catholique de Lille, F-59000 Lille, France. - remi.suchon@univ-catholille.fr

1 Introduction

Cooperation is one of the most important reasons for the success of the human species. However, humans do not always succeed to cooperate. This is especially true in dilemma situations, where narrow self-interest induces agents to refrain from cooperation although mutual cooperation would be beneficial for everyone involved. Such dilemma situations have been studied extensively during the last decades, using theoretical and empirical methods (for an overview, see e.g. Camerer (2011)). It has been suggested that one can overcome such dilemmas by repeated interactions between the agents involved (see the literature on repeated games and on relational contracts, e.g. Brown et al. (2004); Mailath et al. (2006); Dal Bó and Fréchette (2018)). The problem can be also solved if agents are social-minded, i.e. when they care not only about their self-interest, but also about the welfare of everyone involved (see the literature on gift-exchange games (Fehr et al., 1993), public good games, etc). Typically, such a solution to the dilemma problem requires trust. If a social-minded agent trusts that the partners "do the right thing" even if it is not in their narrow self-interest, a social-minded agent will cooperate. But if a social-minded agent does not trust his partners, he will refrain from cooperation to avoid being abused (see e.g. the literature on conditional cooperation in public good games, originated by Fischbacher et al. (2001); Fischbacher and Gächter (2010)).

A lot of economic interactions, in markets or elsewhere, require trust between strangers, where trust is particularly hard to establish. In this case, a mechanism

of information disclosure might help. If one interacts for the first time with another agent, but one is informed that this agent has been trustworthy most of the times in the past in similar situations, one has more trust in this new partner than if the new guy has a reputation for abusing trust.

In theory, this intuition can be supported by at least two complementary mechanisms. First, disclosing one's past record may help sustain indirect reciprocity (Nowak and Sigmund, 2005): A stranger might punish an individual with a record of defection by restraining trust/cooperation, or by refusing to interact with such a person at all. This in turn gives a strong incentive to cooperate. Second, a good record might change the trustor's beliefs about the social-preference type of the trustee and as a consequence help cooperation (Kreps and Wilson, 1982). A lot of electronic market places have implemented mechanisms of information disclosure to improve trust (of the buyer or the employer) and trustworthiness (of the seller or the job candidate). A number of lab experiments have confirmed that information disclosure helps sustaining cooperation in sequential social dilemma (Bolton et al., 2004; Bohnet and Huck, 2004; Charness et al., 2011; Duffy et al., 2013).[1]

In this paper, we investigate experimentally how information on past behavior influences the willingness to trust and to cooperate in dilemma situations with partner choice. In particular, we study how properties of the information revelation mech-

[1] A related literature study the effect of information about choices on cooperation in simultaneous interactions. See for instance Duffy and Ochs (2009); Camera and Casari (2009); Kamei and Putterman (2016).

anism impact its effectiveness on achieving cooperative outcomes. We compare mandatory with voluntary information disclosure mechanisms. Mandatory disclosure of past behavior is quite intrusive and might be at odds with some ethical considerations. In addition, individuals may value privacy and control over their data.[2] For these reasons reputation systems based on voluntary information disclosure may be more desirable. Moreover, most mechanisms have some discretion about revelation of past information. For instance, in online market with information about past behavior, it is often possible for participants to manipulate what to disclose from their past records, for instance by creating a new alias after a history of uncooperative behavior.

In theory, systems with voluntary disclosure should be as effective as mandatory disclosure due to the unravelling principle (Milgrom, 1981). Those who choose not to disclose their record will all be treated the same, so there is an incentive for those with a good record to disclose their records. This leads to unravelling: only those with bad records will withhold information. Since this is anticipated, "non-disclosers" will be treated with skepticism, i.e. one does not cooperate with them or refrain from interacting with them at all. This mechanism provides an incentive to build a good record, i.e. to cooperate. However, evidence from sender-receiver experiments have demonstrated the limits of the unravelling principle. A common observation is that senders with a bad private information exploit receivers who are

[2]See e.g. Varian (2009); Acquisti et al. (2016) for reviews on the economics of privacy and e.g. Benndorf et al. (2015); Benndorf and Normann (2018); Schudy and Utikal (2017) for recent experimental evidence on privacy preferences.

not skeptical enough (see e.g. Jin et al., 2021; Montero and Sheth, 2021; Sheth, 2021, for recent examples).[3] Jin et al. (2021) show that repeated feedback is necessary to reduce the gap between actual behavior and the theoretical prediction of unravelling. In this paper we investigate whether these observed limits of the unravelling principle makes voluntary information disclosure less effective than mandatory information disclosure in promoting cooperation in a sequential dilemma situation with partner choice.

To investigate these questions, we set up a sequential prisoner's dilemma experiment (SPD), where first player 1 (P1 - by convention female) decides whether to cooperate (trust) or defect. After being informed about P1's decision, player 2 (P2 - by convention male) takes his cooperation decision (reciprocate trust or renege). Before the SPD takes place, P1 decides whether the game should be played or whether she takes an outside option instead. The outside option gives both players higher payments than what they get in the SPD if both defect, i.e. if they play the unique Nash equilibrium, but less than what they get when both players cooperate. Hence, if P1 does not trust that P2 would cooperate, she should take the outside option, while if she trusts him, she should opt into the SPD and choose cooperation.

Each subject plays this game repeatedly with fixed roles, but with changing partners with exogenous matching (stranger matching). Before P1 decides whether to pick

[3]There is also evidence of failures of the unravelling principle outside the lab. For instance, Brown et al. (2012) show that film studios exploit the fact that movie goers fail to anticipate that movies which are not reviewed before release tend to be of low quality.

the outside option or the SPD, she might get informed about her prospective P2's past cooperation choices. If information about P2's past choices is disclosed, it is correct and complete. The disclosure of information is either mandatory ("Mand" treatments), or it is voluntary, ie only disclosed if P2 wants so ("Vol" treatments). The information might be disclosed with 100% probability, (no noise, "NN" treatments), or it is disclosed only with 90% probability (low noise, "LN" treatments). This implies four treatments: MandNN, MandLN, VolNN, and VolLN. Besides, in a baseline treatment no information gets disclosed. In this treatment P1 has to choose between the SPD and the outside option without having any information about the previous choices of her prospective P2. The noise treatments allow us to measure the robustness of voluntary disclosure when seeing no information might be plausibly blamed on bad luck. They also allow us to explore skepticism, by comparing the behavior of P1 when lack of information is due to P2's choice (in the VolNN treatment) with the behavior of P1 when the lack of information might be due to chance (in the LN treatments).

As expected, we find that disclosure of the record of P2s' past choices increases P1s' trust levels as measured by their cooperation choices. As a result, full cooperation by both players is significantly more often observed when such an information-disclosure system is in place. The interplay between information-disclosure and partner's choice can mitigate the problem of dilemma situations. In contrast to the existing literature identifying limits of unravelling, voluntary disclosure is just as effective as mandatory disclosure. When given the choice, P2s reveal their records

most of the time, and the probability of information disclosure increases with the quality of P2's record. P1s anticipate this, and they are skeptical: In the voluntary disclosure treatments, they do not trust P2s of whom they do not see the records. As a placebo test, we compare trust by P1s when seeing the records and when not seeing it in the MandLN treatment. In this treatment the impact of information on P1s' cooperation rates is much smaller than in the VolLN treatment. Importantly, we find that the level of disclosure and skepticism are already high in the first periods of the game, suggesting that learning plays only a limited role. Clearly, these results are in contrast with the limits of unravelling found in sender-receiver experiments. In addition, our results also reveal that information disclosure is much less effective when the information-system is not perfectly reliable, ie in treatments with noise. Even the rather low noise levels of the LN-treatments are enough to reduce the overall cooperation level substantially.

We are aware of only two papers comparing automatic and voluntary disclosure in a social dilemma (Kamei, 2017, 2020). Both focus on simultaneous games, while we use a sequential prisoner's dilemma. Sequentiality is common in real life economic interactions, including online transactions (the buyer must pay before the sellers sends the product) or employer-employee relations. In contrast to simultaneous games, in a sequential game all the strategic uncertainty weights on the first mover (as noted by e.g. Ghidoni and Suetens, 2022). In SPD, P2 does not have to form a belief about P1's choice, because P1's choice is already known to P2 when P2 makes his own choice. Hence, any cooperation choice of previous P1s should be irrelevant

for P2's choice in a particular round. This in turn implies that in our game, P2's record reveals all the information P1 might need, while in the simultaneous game P1 would also have to know the cooperation choices of P2's prior partner to interpret P2's record correctly. Therefore, in both Kamei (2017) and Kamei (2020), the information about one's opponent is more complex to interpret. Both papers find that, under random matching, the possibility to freely hide one's identity (Kamei, 2017) or one's last action (Kamei, 2020) undermines the effect of information disclosure on cooperation. In contrast, our results indicate that voluntary disclosure may be just as effective as mandatory one, in a context in which the interpretation of the disclosed information is relatively straightforward.

The remainder of the paper is organized as follows. Section 2 details the experimental design, Section 3 presents our results and Section 4 discusses these results and concludes.

2 Design

2.1 The stage game

The stage game is a modified sequential prisoner's dilemma. Compared to a regular sequential prisoner's dilemma the main modification is that in each round P1 can choose whether she enters the game or not. If she decides not to enter, both players get an outside option with a fixed and equal payoff. The modified sequential prisoner's dilemma is shown in Figure 1.

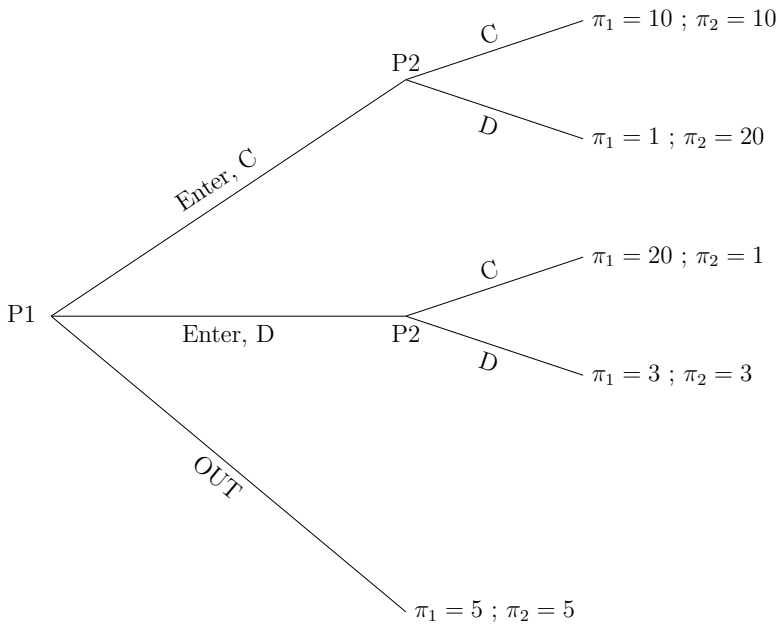


Figure 1: The stage game.

P2 plays the strategy method: he takes one decision for the node after P1 decided to enter and defect, and one decision for the node after P1 decided to enter and cooperate. These two choices are made before P2 gets informed about P1's actual choice. The use of the strategy method allows to gather more data and to insure that the history of play of a P2 does not depend on her past partners' behavior. The use of strategy method to elicit P2s' choices is common and does not affect behavior (Brandts and Charness, 2000, 2011).

The parameters were chosen to make cooperation hard to sustain: the temptation to defect for P2 is high, which is expected to deter entry and cooperation of P1 (e.g. Mengel, 2017; Gaechter et al., 2022, study the impact of game parameters on cooperation in the prisoner's dilemma).

2.2 Matching

The stage game is played for 30 periods, which is common knowledge. Roles are fixed: each player is randomly assigned to be P1 or P2, once and for all before the first round. At every round, each P1 is matched with a randomly selected P2 (stranger matching).[4] Participants are payed for every period.

2.3 Experimental treatments

There are 5 experimental treatments. The **Baseline treatment** is fully described by the stage game and the matching introduced earlier. Beside the **Baseline treatment**, we have 4 treatments in which information about P2 may be revealed to P1 before she decides to enter. We call this information *the record* of P2. The record contains two pieces of information about P2. First, it reports the portion of times that P2 chose "Cooperate" at the Enter Cooperate node. Second, it reports the portion of times that P2 chose "Cooperate" at the Enter Defect node. Remember that P2s' choices are elicited using the strategy method. Therefore, the observed decision of a P2 in a particular period is independent of the actual choice of her P1 in this period and the record of a P2 is independent of the actual choice of the P1 he was matched with in past periods. Figure A.1 in Appendix A gives an example of a record as disclosed to P1.

[4]Given the size of our sessions, participants meet more than once on average. This could in principle affect the evolution of cooperation through contagion. Ghidoni et al. (2019) show that random matching in groups of 6 participants give similar results to perfect stranger's matching. In all our sessions we had more than 6 participants. In addition, all our results hold when we control for the number of participants in the respective session.

Record disclosure is:	Mandatory	Voluntary
Certain	MandNN	VolNN
Noisy	MandLN	VolLN

Table 1: Summary of the treatments with record disclosure.

The 4 information treatments differ in the way the record is disclosed to P1. There are two dimensions: First, record disclosure might be mandatory (the "Mand" treatments), or it might be P2's choice whether to disclose his record or not (the "Vol" treatments). Second, we vary whether disclosure is noisy: in the LN treatments, there is a 10% chance that the record is not transmitted to P1 when information transmission is mandatory (in the MandLN treatment), or when P2 chose to reveal his record (in the VolLN treatment). In the NN treatments, no such noise is introduced. These variations result in 5 treatments: **Baseline, MandNN, MandLN, VolNN, VolLN**. Table 1 summarizes the information treatments.

2.4 Procedure

We recruited a total of 386 participants for 25 sessions.[5] 14 sessions were run at the BEEL lab in Brussels in winter 2019/2020 (before the pandemic), while 11 sessions were run at the Anthropo-lab in Lille in fall 2021 (after the pandemic).

Table 2 sums up the distribution of participants and sessions across treatments and cities. Our results are robust to the inclusion of a dummy variable indicating the

[5]A preregistration of the experiment can be found at https://doi.org/10.1257/rct.4937. Note that, due to constraints related to the pandemic, we had to drop two treatments in which a larger noise was introduced. We have a power of 80% to detect an effect of 10 percentage points at the 5% level (more details are in Appendix B). This effect size is less than half the size of the effect of disclosing the second movers' history reported in Charness et al. (2011).

	N Sessions (in Lille)	N Participants (in Lille)
Baseline	5 (3)	84 (52)
MandNN	5 (2)	74 (38)
VolNN	5 (2)	84 (42)
MandLN	5 (3)	76 (56)
VolLN	5 (1)	68 (20)
Total	25 (11)	386 (208)

Number in parentheses are for the sessions run in Lille.

Table 2: Summary of the sessions.

city in which the session was run.

On average a session lasted about 1 hour. The average earnings were €12.87 (S.D.:2.37). Before starting the experiment, participants had to successfully pass a non-incentivized understanding questionnaire. At the end of the experiment, participants had to fulfill a demographic questionnaire. The full set of instructions can be found in Appendix F.

3 Results

Table 3 reports the descriptive statistics by treatment.[6]

The cooperative outcome is relatively rarely observed. This was expected because our parameters make defection tempting. P1s enter in the vast majority of the cases, but cooperate much less often. P2s cooperate much more often following cooperation, than following defection, but still cooperate following defection by P1 in

[6]In Table D.1 in Appendix D.1, we report the dynamics of the Cooperative outcome for each session separately.

Treatment	Coop. outcome	P1 enters	P1 coop.	P2 coop. if:		P2 discloses
				P1 coop.	P1 def.	
Baseline	0.110	0.686	0.289	0.358	0.179	-
MandNN	0.264***	0.801**	0.420***	0.533**	0.159	-
VolNN	0.236***	0.769*	0.367**	0.530**	0.140	0.619
MandLN	0.157	0.740	0.326	0.365	0.163	-
VolLN	0.213*	0.742	0.334	0.473	0.140	0.556

Test vs Baseline from Logit regressions with standard errors clustered at the session level. $*p < 0.10, **p < 0.05, ***p < 0.01$. Table C.1 in Appendix C provides details.

Table 3: Descriptive statistics

roughly 15% of the cases. One may fear that irrationality plays a big role in both "over-entry" by P1s and cooperation of P2s following defection by P1. A careful analysis of this question is reported in Appendix E, the result of which is that irrationality is not a big concern.

The time dynamic of the cooperative outcome is not impacted by the presence of a reputation system. There is a decline in the likelihood of the cooperative outcome, but it is small, and similar with or without record disclosure. See Appendix D.2 for more details.

3.1 Cooperation

Result 1: The existence of a record system increases the likelihood of the cooperative outcome. Voluntary record disclosure is as effective as mandatory disclosure. On the other hand, introducing a small noise decreases the effectiveness of record disclosure substantially.

Support for Result 1: The cooperative outcome occurs in 11% of cases in the baseline, and in 21.7% of cases in the other treatments ($p < 0.01$, see Table C.2 and C.3 in Appendix C for details). To disentangle the respective effect of voluntary disclosure and noise, we run logit regressions explaining the occurrence of the cooperative outcome by dummy variables indicating a voluntary disclosure system and a noisy disclosure system, and the interaction of these two dummies. We control for the presence of a record system in every models. The respective marginal effects of voluntary and noisy disclosure are reported in Table 4.

	(1) Coop. out.=1	(2) Coop. out.=1	(3) Coop. out.=1
Marginal Effect: (at Record=1)			
Voluntary dis.	0.006	0.006	0.001
	(0.031)	(0.031)	(0.032)
Noisy dis.	-0.061**	-0.061**	-0.062**
	(0.029)	(0.029)	(0.028)
Observations	5780	5780	5780
Session char.	No	No	Yes
Period FE	No	Yes	Yes
Sessions	25	25	25

Standard errors in parentheses are clustered at the session level.
$*p < 0.10, **p < 0.05, ***p < 0.01$
Session characteristics: City dummy and size of the session.

Table 4: The marginal effects of voluntary and noisy disclosure.

Since we used the strategy method, we observe the four pure strategies of P2s: i) Cooperate when P1 cooperated (conditional cooperation), ii) Cooperate irrespective of the decision of P1 (unconditional cooperation), iii) Never cooperate (unconditional defection) and iv) Cooperate only when P1 defected (mismatch). In what follows,

to study cooperation choices of P2, we pool together conditional and unconditional cooperation. Focusing strictly on conditional cooperation does not change our results. More details about the distribution of strategies across treatments are given in Appendix D.3.

Result 2: When there is no noise, the existence of a record system increases cooperation levels of the P1s and P2s irrespective of whether disclosure is mandatory or voluntary. With noise the record system has no significant effect on cooperation. The effect of a record system on the enter decision is small and at best weakly significant.

Support for Result 2: Table 3 provides strong first evidence in favor of Result 2. To disentangle the respective effects of noise and voluntary disclosure we run random-effect logit regressions explaining the occurrence of the entering, and cooperation of P1 and P2 by dummy variables indicating a voluntary disclosure system and a noisy disclosure system and the interaction of these two dummies. The marginal effects of "noisy" and "voluntary" are reported in Table 5. We control for the presence of a record system in every models. The effect of voluntary disclosure is never significant. The effect of noise on entry is not significant, but it is significant on P1s' and P2s' cooperation levels: noise reduces P1s' cooperation by more than 7 percentage points (p=0.018) and P2s' cooperation by more than 11 pp (p=0.011).

	(1) Enter=1	(2) P1's Coop=1	(3) P2's Cond Coop=1
Marginal effect of:			
Voluntary	-0.014	-0.036	0.031
	(0.038)	(0.031)	(0.055)
Noisy	-0.056	-0.073**	-0.112**
	(0.044)	(0.031)	(0.044)
Observations	5780	5780	5780
Session characteristics	Yes	Yes	Yes
Individual characteristics	Yes	Yes	Yes
Period FE	Yes	Yes	Yes
Sessions	25	25	25

Standard errors in parentheses are clustered at the session level.

* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001

Session characteristics: City dummy and size of the session.

Individual characteristics: Gender, age, occupational status, experience with experiments.

Table 5: The effects of voluntary disclosure and of noise

3.2 Disclosure, Record and skepticism

Noise has no effect on the disclosure decisions of the P2s. As one can see from Table 3 the disclosure rates are very similar in both voluntary disclosure treatments. When disclosure is voluntary, the disclosure decision depends mainly on P2's record.

Result 3: P2s with better records are more likely to disclose it. This is true for both the VolLN and the VolNN treatment.

Support for Result 3: We estimated random-effect logit models explaining the decision of a P2 to disclose his record by the rate at which he cooperated following cooperation by P1 and following defection in the past. Obviously, we use only the data from the Vol treatments. Table 6 reports the outcome of these regressions. Column (1) reports the marginal effects pooling the data from both Vol treatments. Column (2) contrasts the marginal effects of the VolNN (Noisy=0) and

VolLN (Noisy=1) treatments. The probability of disclosure obviously increases in the level of cooperation following cooperation of P1 in the past. The effect of past cooperation following a defection of P1 is very small and not significant. Again, we do not see any significant impact of noise on the disclosure decision.

	(1) disclose=1	(2) disclose=1
Marginal Effect:		
% of Coop. when P1 cooperated	0.542****	
	(0.114)	
% of Coop. when P1 defected	0.047	
	(0.100)	
Marginal Effect:		
% of Coop. when P1 cooperated at:		
Noisy=0		0.705****
		(0.074)
Noisy=1		0.376**
		(0.173)
% of Coop. when P1 defected at:		
Noisy=0		0.024
		(0.140)
Noisy=1		0.01
		(0.151)
Observations	2204	2204
Session characteristics	Yes	Yes
Individual characteristics	Yes	Yes
RE/FE	RE	RE
SE	Cluster	Cluster

Standard errors in parentheses are clustered at the session level.
* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001
Session characteristics: City dummy and size of the session.
Individual characteristics: Gender, age, occupational status, experience with experiments.

Table 6: The marginal effect of one's record on the probability of disclosure

This result can be also seen from Figure 2 that shows the predicted probability of disclosure depending on portion of P2's cooperation choices in all the past rounds.

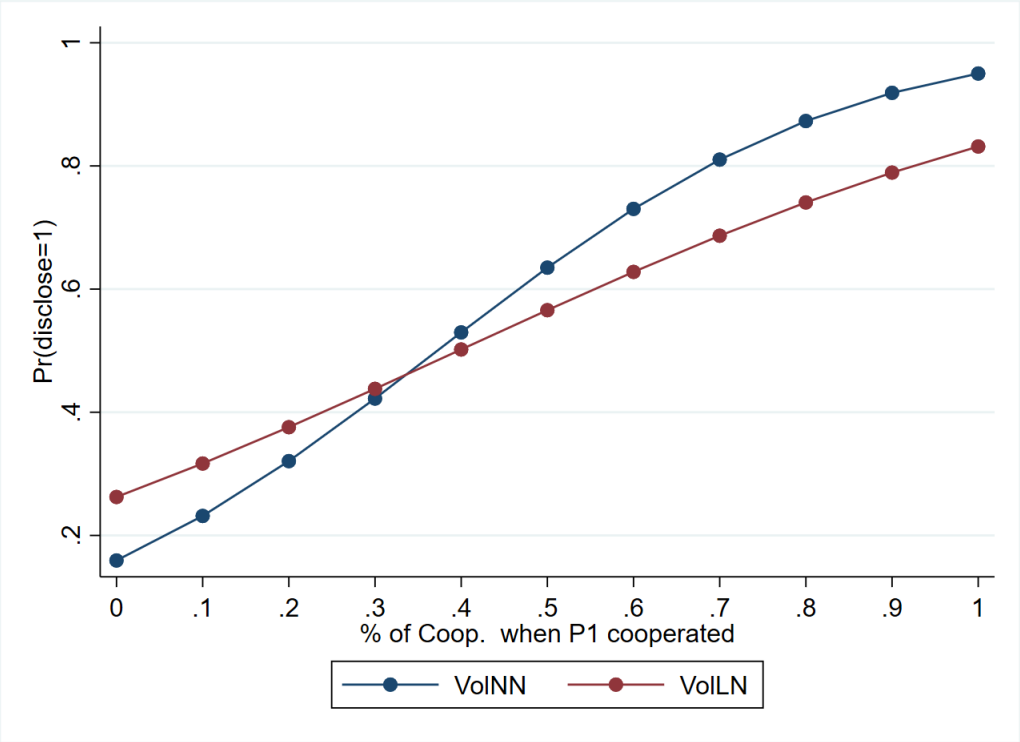


Figure 2: Predicted probability to disclose depending on one's past record.

Result 4: P1s cooperate more when they see the record of P2. In addition, a record indicating that P2 cooperated more in the past, irrespective of whether this follows cooperation by P1 or not, increases P1's likelihood to enter. In contrast, a record of more cooperation of P2 following cooperation of P1 increases P1s' likelihood to cooperate, while a record of more cooperation after defection leads to less cooperation of P1.

Support for Result 4: Figure 3 shows the cooperation of P1s depending on whether P1s saw the record of P2, separating treatments with mandatory and voluntary disclosure. Overall, seeing the record increases trust, i.e P1s' cooperation rates, by about 16 percentage points pooling all the data from the different treatments ($p < 0.001$, see Table C.4 in Appendix C.3 for details on the tests). The

positive effect of seeing the record is seen in all three treatments where the disclosure is neither excluded nor guaranteed by design.

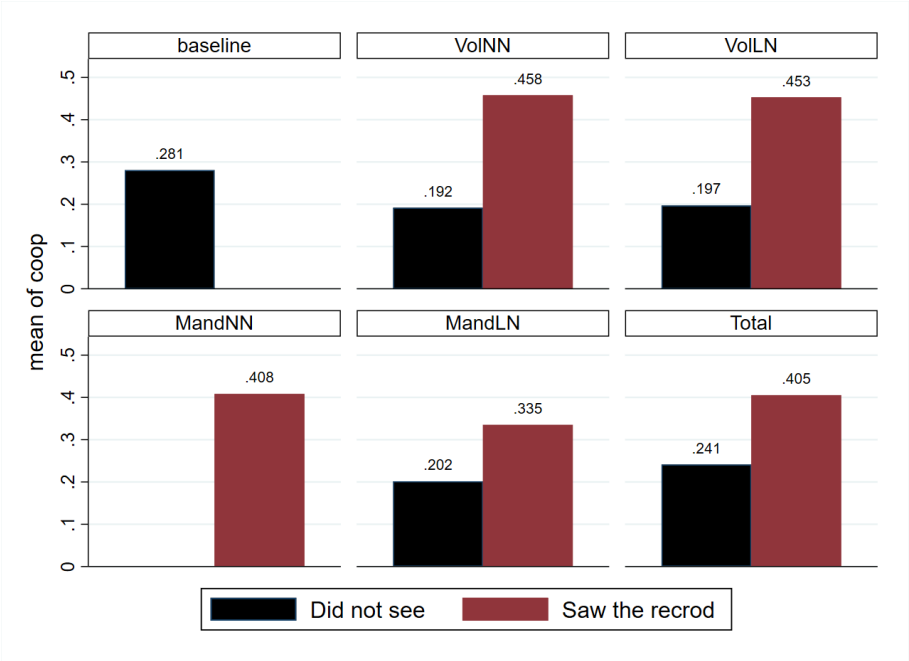


Figure 3: Cooperation rate depending on whether P1 saw the record, separate by voluntary / mandatory disclosure.

We run logit regressions on the subset of the data in which the record of P2 is displayed to P1. We explain the decision of P1 to cooperate depending on the record of P2, i.e. on the portion of times P2 chose cooperation in the past. We present the outcome of several models in Table 7, including random and fixed effects models. For all specifications, the estimated parameters for P2s' previous cooperation are significant at the 1% level. Note that, in contrast with all other parameters, the sign for the effect of previous cooperation following defection on cooperation by P1 is negative. This make intuitive sense: the more likely P2 is to cooperate even after P1s' defection, the more likely P1 is to enter and defect.

	(1) enter=1	(2) enter=1	(3) coop=1	(4) coop=1
Marginal effect of record:				
% of Coop. of P2 when P1 cooperated	0.389****	0.366****	0.513****	0.452****
	(0.053)	(0.019)	(0.037)	(0.025)
% of Coop. of P2 when P1 defected	0.190****	0.221****	-0.257****	-0.268****
	(0.033)	(0.039)	(0.052)	(0.051)
Observations	3365	2758	3382	3284
Session characteristics	Yes	-	Yes	-
Individual characteristics	Yes	-	Yes	-
p-value diff.	< 0.01	< 0.01	< 0.01	< 0.01
RE/FE	RE	FE	RE	FE
SE	Cluster	Boot.	Cluster	Boot.

Standard errors in parentheses are clustered at the session level.

* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001

Session characteristics: City dummy and size of the session.

Individual characteristics: Gender, age, occupational status, experience with experiments.

Table 7: The effect of the record of P2 on the decision to enter and cooperate by P1

To illustrate this result, Figure 4 reports the predicted probability of cooperation by P1, depending on the record of P2, for each treatment with information disclosure (confidence interval are omitted for readability). It shows that the better the record, the higher the probability of cooperation by P1. This does not change across treatments.

Result 5: P1s show skepticism: In the Vol treatments P1s cooperate less when they do not see the record of P1 than when they do see it. This is not the case in the MandLN treatment. Skepticism is observed already in the early rounds of the game, and at best limited learning takes place.

Support for Result 5: We already checked that P1s are more likely to cooperate when they see the report of P2. We expect that the impact of non-disclosure on P1s'

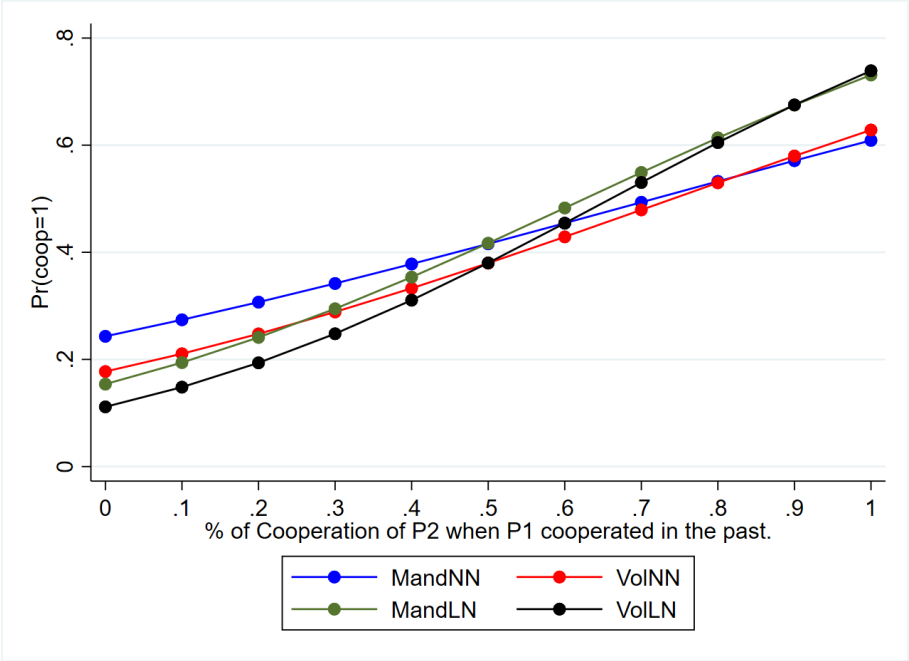


Figure 4: The predicted probability of cooperation by P1 depending on the record of P2.

cooperation is weaker when non-disclosure cannot be attributed to a deliberate decision of P2, since in this case non-disclosure cannot serve as a signal of a poor record.

As one can see from Figure 5 non-disclosure reduces the P1s' cooperation rates by 10 percentage points (from 37.3 to 27.3%) in the Mand treatments where non-disclosure cannot be voluntary. In the Vol treatments, where non-disclosure is mainly due to deliberate decisions of the P2s, the non-disclosure induced a drop in P1s' cooperation rates more than twice as large, namely 26.2 percentage points (from 45.6 to 19.%). This is a first indication of the validity of Result 5. We run a random-effect logit model explaining P1's decision to cooperate by a dummy variable indicating voluntary disclosure and a dummy variable indicating that P1 saw the record, and an interaction of these two dummies. We test whether the marginal effect of seeing

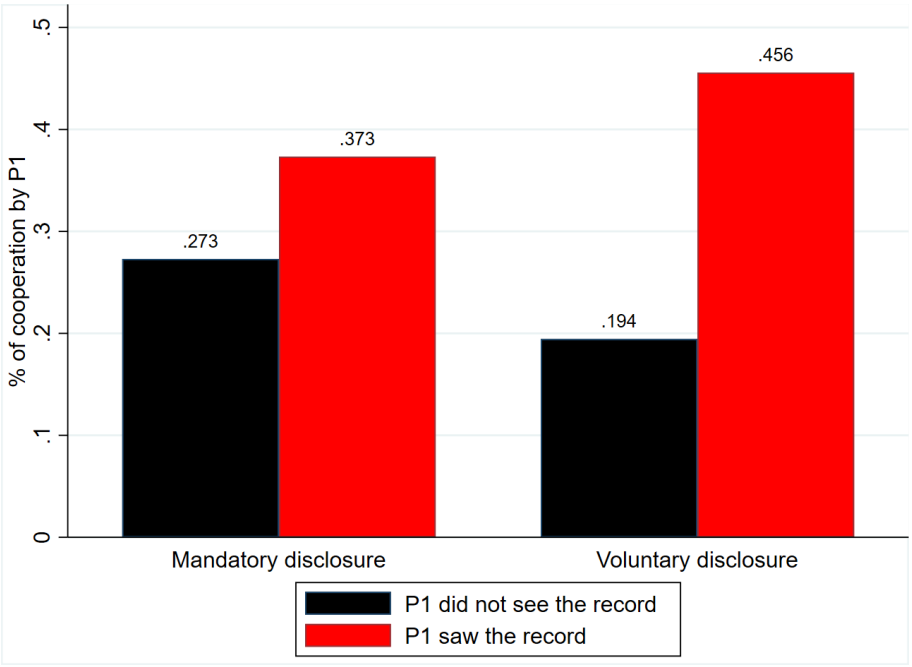


Figure 5: Cooperation of P1 depending on whether P1 saw the record, contrasted by voluntary / mandatory disclosure.

the record is different when disclosure is voluntary and when it's not. Table 8 shows the outcomes and confirms the result suggested by Figure 5: the effect of (not) seeing the record is much weaker when disclosure is for sure not voluntary (p=0.01) . In addition, in Column 2, we interact our variables of interest with a categorical variable indicating the period bracket. The values of the interactions suggest that the effect of not seeing the record is already present early in the game, and increases moderately.

4 Discussion and Conclusion

We studied the effect of different information-disclosure mechanisms on cooperation in a sequential social dilemma. We found that such institutions increase the likelihood of reaching the cooperative outcome, as both the first and second movers act

	(1) Coop.=1	(2) Coop.=1	(3) Coop.=1
Marginal effect of seeing the record at:			
Voluntary=0	0.137****		
	(0.035)		
Voluntary=1	0.256****		
	(0.030)		
p-value diff.	0.010		
Voluntary=0 × Period≤ 10		0.074**	
		(0.036)	
Voluntary=0 × 10 <Period≤ 20		0.197****	
		(0.043)	
Voluntary=0 × Period> 20		0.138***	
		(0.048)	
Voluntary=1 × Period≤ 10		0.202****	
		(0.035)	
Voluntary=1 × 10 <Period≤ 20		0.282****	
		(0.028)	
Voluntary=1 × Period> 20		0.284****	
		(0.049)	
VolNN			0.279****
			(0.043)
MandLN			0.144***
			(0.046)
VolLN			0.231****
			(0.038)
Observations	5587	5587	5587
RE/FE	RE	RE	RE
SE	Cluster	Cluster	Cluster
Session Char.	Yes	Yes	Yes
Indiv. Char.	Yes	Yes	Yes

Standard errors in parentheses are clustered at the session level.

* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001

Session characteristics: City dummy and size of the session.

Individual characteristics: Gender, age, occupational status, experience with experiments.

Table 8: The differential effect of (not) seeing the record, depending on the treatment.

more cooperatively. In addition, we found that an institution in which disclosure is voluntary is exactly as effective as one in which it is mandatory. In contrast, we found that a relatively small noise significantly undermines the effect of information

disclosure.

A striking result is that voluntary disclosure is exactly as effective as mandatory disclosure. This result is consistent with unravelling, but it is in contrast with two previous experiment comparing voluntary and automatic disclosure of information in social dilemma Kamei (2017, 2020). There are several differences between these papers and ours that may explain the different results. First, the underlying social dilemma are simultaneous in these papers, while it is sequential in ours. This changes the distribution of strategic uncertainty between the first and the second mover (Ghidoni and Suetens, 2022), which may impact the effectiveness of information disclosure. Second, and more importantly, the information disclosed in the work of Kamei is more complex. In the simultaneous games, the second order information is not strictly controlled for, so the interpretation of one's counterpart's record depends on the belief one has on the records of the individuals with whom the counterparts interacted in the past. This may induce the (maybe mistaken) beliefs that a bad record will not be sanctioned with (a high likelihood of) defection. In our experiment, sequentiality, combined with the strategy method, makes the interpretation of one's record much more straightforward, which might foster the incentives to behave as a cooperative type.

More generally, our results also contrast with the literature on the failure of unravelling in sender-receiver games. In this literature, senders typically send favorable information, but withhold unfavorable ones, exploiting the naivety of (some of) the

receivers. This failure of unravelling is corrected only after some learning has occurred. This tends to suggest that voluntary disclosure should be significantly less effective than mandatory one, if effective at all. In our experiment, when given the choice, a large share of the subjects choose to maintain a good record and disclose it (see Figure D.7 in Appendix D.4.). First movers are skeptical: they generally avoid exposing themselves to being exploited by a second mover who does not disclose his record. This is observed already early in the game, and learning is limited. What can explain this difference between our results and that of sender-receiver game experiments? A first possibility relates to the *nature* of the information disclosed. In sender-receiver games, participants disclose a random variable which is exogenously determined and contains no signal about the senders' type. In contrast, building and sustaining a good reputation may have an intrinsic value for a participant in social dilemma game, because a good record signals to oneself as well as to others that one is a cooperative, "nice" type (see e.g. Bénabou and Tirole, 2006, and the large subsequent literature on identity management). Note also that P2s' reputation in our experiment depends on all his previous choices. Hence, a P2 cannot hope to get a "clean" record after he has refused to cooperate with a cooperating P1. In contrast, in the sender-receiver game, the game starts from "scratch" in every round. Because of the reputational spillover between rounds participants might experiment less in our SPD than in a sender-receiver game. Irrespective of the fundamental reasons, our results suggest that the limits of unravelling identified in past sender-receiver experiments might depend on the nature of the strategic situation.

Another striking result is the effect of noise on the effectiveness of information disclosure. We designed the treatment with noise with two goals in mind: assessing the robustness of the effect of information disclosure, and providing a placebo test for skepticism (comparing the behavior of P1 when not seeing the record can vs cannot be blamed on luck). The strong effect of the relatively small noise may be explained by the often observed tendency of individuals to overweight small probabilities (Tversky and Kahneman, 1992).

It would be interesting to see whether agents prefer mandatory disclosure or voluntary disclosure mechanisms in dilemma situations. On the one hand, the experimental results suggest that mandatory disclosure is never worse and in simultaneous games even better than voluntary disclosure. This should induce agents to opt for a mandatory disclosure mechanism for general dilemma situations. On the other hand, it is well know that humans like (the illusion of) control, and this motivation is of course more in line with voluntary disclosure.

References

Acquisti, A., C. Taylor, and L. Wagman (2016, June). The economics of privacy. *Journal of Economic Literature 54*(2), 442–92.

Benndorf, V., D. Kübler, and H.-T. Normann (2015). Privacy concerns, voluntary disclosure of information, and unraveling: An experiment. *European Economic Review 75*, 43–59.

Benndorf, V. and H.-T. Normann (2018). The willingness to sell personal data. *The Scandinavian Journal of Economics 120*(4), 1260–1278.

Bohnet, I. and S. Huck (2004, May). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review 94*(2), 362–366.

Bolton, G. E., E. Katok, and A. Ockenfels (2004). How effective are electronic reputation mechanisms? an experimental investigation. *Management Science 50*(11), 1587–1602.

Brandts, J. and G. Charness (2000). Hot vs. cold: Sequential responses and preference stability in experimental games. *Experimental Economics 2*(3), 227–238.

Brandts, J. and G. Charness (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics 14*(3), 375–398.

Brown, A. L., C. F. Camerer, and D. Lovallo (2012, May). To review or not to

review? limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics 4*(2), 1–26.

Brown, M., A. Falk, and E. Fehr (2004). Relational contracts and the nature of market interactions. *Econometrica 72*(3), 747–780.

Bénabou, R. and J. Tirole (2006, December). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Camera, G. and M. Casari (2009, June). Cooperation among strangers under the shadow of the future. *American Economic Review 99*(3), 979–1005.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Charness, G., N. Du, and C.-L. Yang (2011). Trust and trustworthiness reputations in an investment game. *Games and Economic Behavior 72*(2), 361–375.

Dal Bó, P. and G. R. Fréchette (2018, March). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature 56*(1), 60–114.

Duffy, J. and J. Ochs (2009). Cooperative behavior and the frequency of social interaction. *Games and Economic Behavior 66*(2), 785–812.

Duffy, J., H. Xie, and Y.-J. Lee (2013). Social norms, information, and trust among strangers: Theory and evidence. *Economic theory 52*(2), 669–708.

Engelmann, D. and M. Strobel (2004, September). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics 108*(2), 437–459.

Fischbacher, U. and S. Gächter (2010, March). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review 100*(1), 541–56.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Gaechter, S., K. Lee, M. Sefton, et al. (2022). The variability of conditional cooperation in sequential prisoner' s dilemmas. Technical report.

Ghidoni, R., B. L. Cleave, and S. Suetens (2019). Perfect and imperfect strangers in social dilemmas. *European Economic Review 116*, 148–159.

Ghidoni, R. and S. Suetens (2022). The effect of sequentiality on cooperation in repeated games. *American Economic Journal: Microeconomics*.

Jin, G. Z., M. Luca, and D. Martin (2021, May). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics 13*(2), 141–73.

Kamei, K. (2017). Endogenous reputation formation under the shadow of the future. *Journal of Economic Behavior & Organization 142*, 189–204.

Kamei, K. (2020). Voluntary disclosure of information and cooperation in simultaneous-move economic interactions. *Journal of Economic Behavior & Organization 171*, 234–246.

Kamei, K. and L. Putterman (2016, 07). Play It Again: Partner Choice, Reputation Building and Learning From Finitely Repeated Dilemma Games. *The Economic Journal 127*(602), 1069–1095.

Kreps, D. M. and R. Wilson (1982). Reputation and imperfect information. *Journal of Economic Theory 27*(2), 253–279.

Mailath, G. J., L. Samuelson, et al. (2006). *Repeated games and reputations: long-run relationships*. Oxford university press.

Mengel, F. (2017, 12). Risk and temptation: A meta-study on prisoner's dilemma games. *The Economic Journal 128*(616), 3182–3209.

Miettinen, T., M. Kosfeld, E. Fehr, and J. Weibull (2020). Revealed preferences in a sequential prisoners' dilemma: A horse-race between six utility functions. *Journal of Economic Behavior & Organization 173*, 1–25.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380–391.

Montero, M. and J. D. Sheth (2021). Naivety about hidden information: An experimental investigation. *Journal of Economic Behavior & Organization 192*, 92–116.

Nowak, M. A. and K. Sigmund (2005). Evolution of indirect reciprocity. *Nature 437*(7063), 1291–1298.

Schudy, S. and V. Utikal (2017). 'you must not know about me'—on the willingness to share personal data. *Journal of Economic Behavior & Organization 141*, 1–13.

Sheth, J. D. (2021). Disclosure of information under competition: An experimental study. *Games and Economic Behavior 129*, 158–180.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty 5*(4), 297–323.

Varian, H. R. (2009). *Economic Aspects of Personal Privacy*, pp. 101–109. Boston, MA: Springer US.

A Screenshots

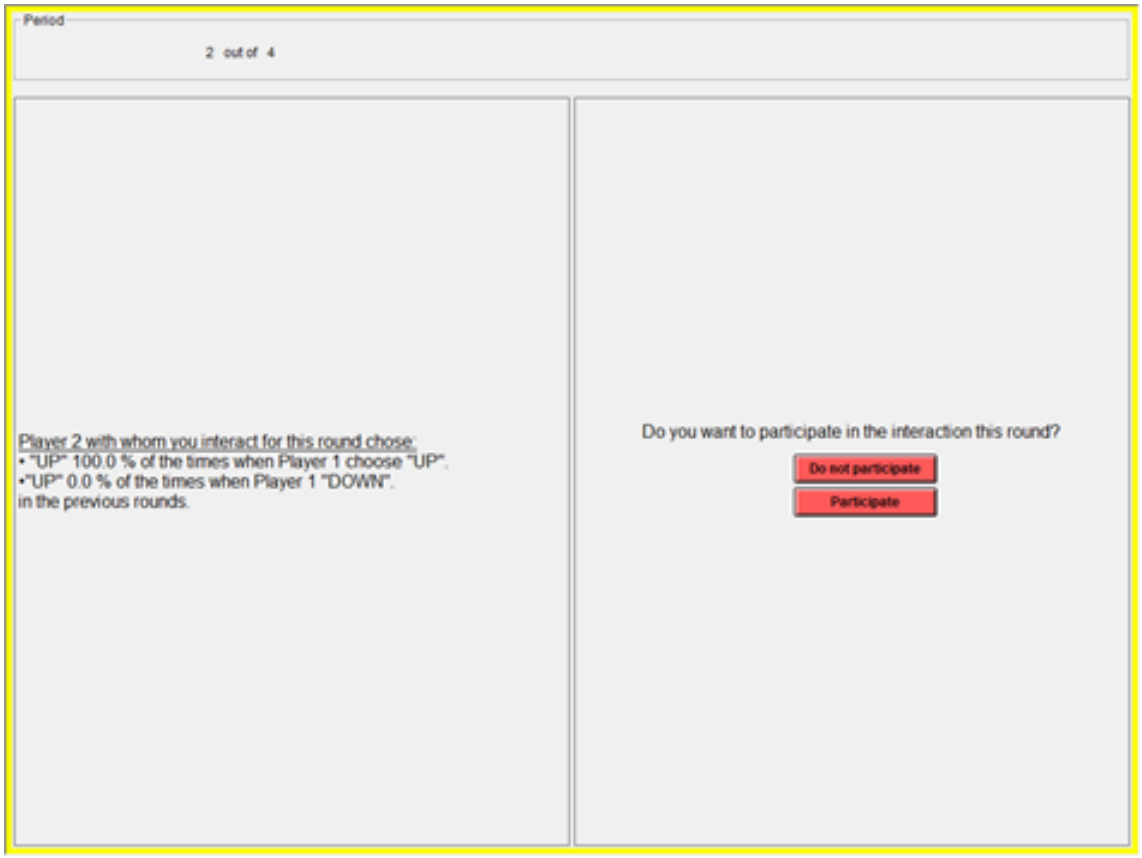


Figure A.1: Record of P2, as displayed to P1

B Power analysis

We followed a simulation approach for our power analysis and sample size determination. The advantage of simulation based power analysis is that we can adapt it to the test we will use on the actual data. We focused on the likelihood of the cooperative outcome as our variable of interest, and proceeded as follows:

- We randomly created samples fixing a number of parameters:
  - The number of observations per session.
  - The number of sessions.
  - The baseline probability of the cooperative outcome.
  - The effect of the existence of an information disclosure mechanism (i.e. an increase in the probability of the cooperative outcome, measured in percentage point).
  - The session effect size (also measured in percentage point).

In each sample, half the observations were allocated to the baseline, and half to a treatment with information disclosure. We generated 100 samples for each parameter space, and for each sample, we ran a logit model explaining the probability of the cooperative outcome by the treatment dummy, with errors clustered at the session level. For each parameter space, we recorded the percentage of the times the treatment dummy was significant at the 5% level. This gives us the power of our experiment for the parameter space. We reproduced it for numerous parameter spaces. In Table B.1, we report the power analysis for some parameters. This

suggests that, if we expect an effect size of 10 pp with a moderate session effect, and session of, on average, 14 participants, we have appropriate power to detect effect size of 10 or 15 percentage points.

Power	Baseline likelihood	Treatment effect	Session effect	N sessions	N participants
.99	.2	.1	.05	10	140
.64	.2	.1	.1	10	140
1	.2	.15	.05	10	140
.96	.2	.15	.1	10	140
.58	.2	.15	.15	10	140

Table B.1: Power for some parameters.

C Support for the results (omitted in the text).

C.1 Test for Table 3

	(1) Coop. out.=1	(2) P1 enters=1	(3) P1 coops=1	(4) P2 coops when P1 coops=1	(5) P2 coops when P1 def.=1
baseline	-	-	-	-	-
MandNN	0.149*** (0.047)	0.113** (0.054)	0.141*** (0.044)	0.156** (0.074)	-0.005 (0.040)
VolNN	0.124*** (0.040)	0.081* (0.042)	0.078** (0.034)	0.168** (0.079)	-0.010 (0.024)
MandLN	0.050 (0.042)	0.031 (0.070)	0.055 (0.049)	-0.000 (0.068)	0.027 (0.035)
VolLN	0.087* (0.046)	0.076 (0.054)	0.049 (0.050)	0.096 (0.082)	0.004 (0.033)
Observations	5780	5780	5780	5780	5780
Session caracteristics	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes
Sessions	25	25	25	25	25

Marginal effects from Logit models.
Standard errors in parentheses are clustered at the session level.
$*p < 0.10$, $**p < 0.05$, $***p < 0.01$, $****p < 0.001$
Session characteristics: City dummy and size of the session.
Individual characteristics: Gender, age, occupational status, experience with experiments.

Table C.1: Regressions for the significance levels reported in Table 3

C.2 Support for Result 1.

Table C.2 reports descriptive statistics. In the second row, the data from all the treatments with disclosure are pooled.

Treatment	Coop. outcome	P1 enters	P1 coop.	P2 coop. if:	
				P1 coop.	P1 def.
Baseline	0.110	0.686	0.289	0.358	0.179
Disclosure[1]	0.217	0.76	0.362	0.476	0.150

1: data from MandNN, MandLN, VolNN & VolLN pooled.

Table C.2: Descriptive statistics, separating baseline and treatments with reputation.

We regress a dummy variable indicating the cooperative outcome on a dummy variable indicating that there is a disclosure system (all treatments pooled). We report the marginal effect from a logit models. Standard errors are clustered at the session level. Results are in Table C.3.

	(1) Coop. outcome=1	(2) Coop. outcome=1	(3) Coop. outcome=1
Disclosure	0.126*** (0.048)	0.126*** (0.047)	0.122*** (0.046)
Observations	5780	5780	5780
Session characteristics	No	No	Yes
Period FE	No	Yes	Yes
Sessions	25	25	25

Standard errors in parentheses are clustered at the session level.

* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001.

Session characteristics: City dummy and size of the session.

Individual characteristics: Gender, age, occupational status, experience with experiments.

Table C.3: The effect of the existence of a disclosure system on the likelihood of the cooperative outcome .

C.3 Support for Result 4.

	(1) Enter=1	(2) Enter=1	(3) Enter=1	(4) Coop.=1	(5) Coop.=1	(6) Coop.=1
Marginal effect of:						
Information disclosed	0.140****	0.211****		0.220****	0.262****	
	(0.020)	(0.026)		(0.025)	(0.021)	
VolNN			0.148****			0.279****
			(0.025)			(0.043)
MandLN			0.054			0.144***
			(0.036)			(0.046)
VolLN			0.196****			0.231****
			(0.028)			(0.038)
Observations	5587	4804	5587	5587	5385	5587
R^2						
RE/FE	RE	FE	RE	RE	FE	RE
SE	Cluster	Bootstrap	Cluster	Cluster	Bootstrap	Cluster
Session Char.	Yes	Yes	Yes	Yes	Yes	Yes
Indiv. Char.	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses are clustered at the session level.

* p<0.10, ** p<0.05, *** p<0.01, **** p<0.001.

Session characteristics: City dummy and size of the session.

Individual characteristics: Gender, age, occupational status, experience with experiments.

Table C.4: The effect of seeing the information about P2's past choices on P1s' choices. Marginal effects from Logit models.

D Additional results and robustness checks

D.1 Results at the session level.

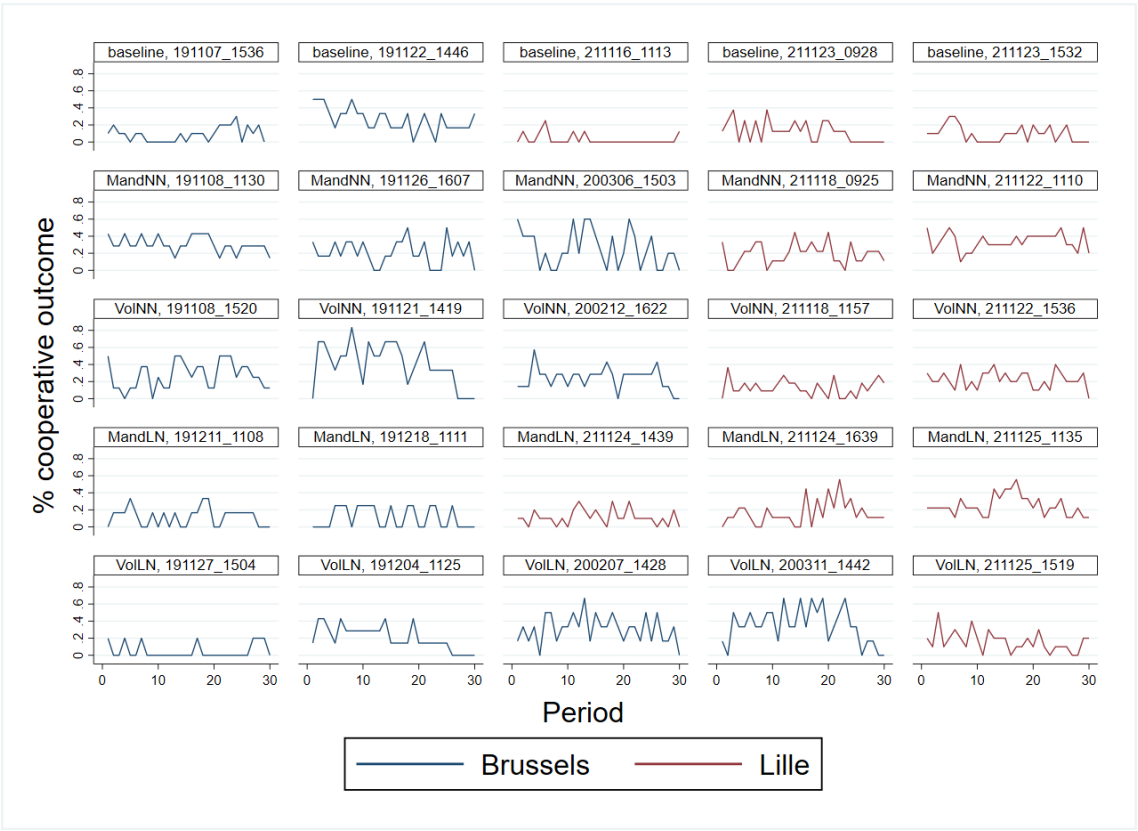


Figure D.1: Cooperative outcomes across periods, for each session separately.

D.2 The time dynamic of entry, (conditional) cooperation, and disclosure

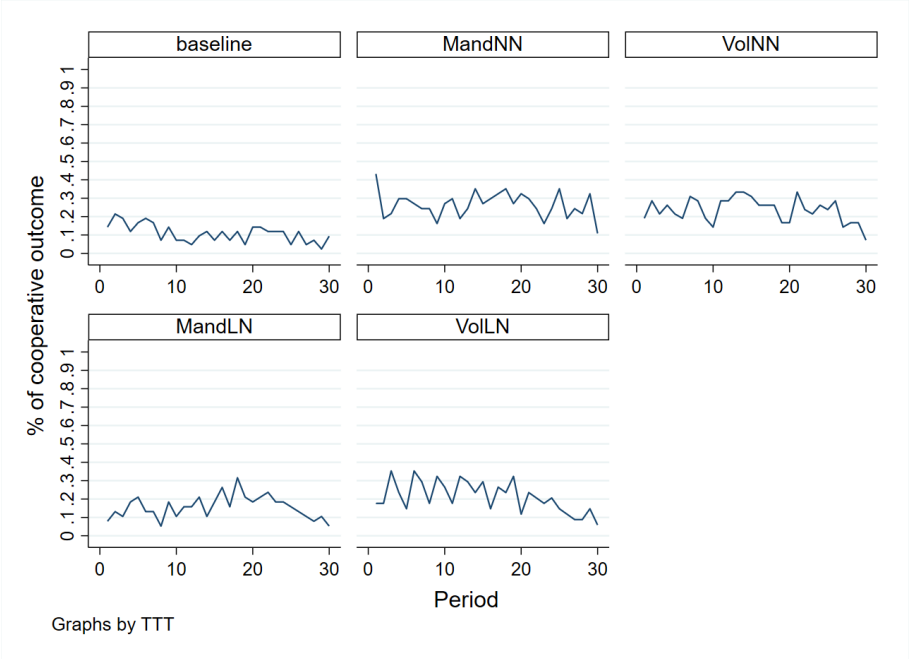


Figure D.2: The dynamic of the cooperative outcome (period averages)

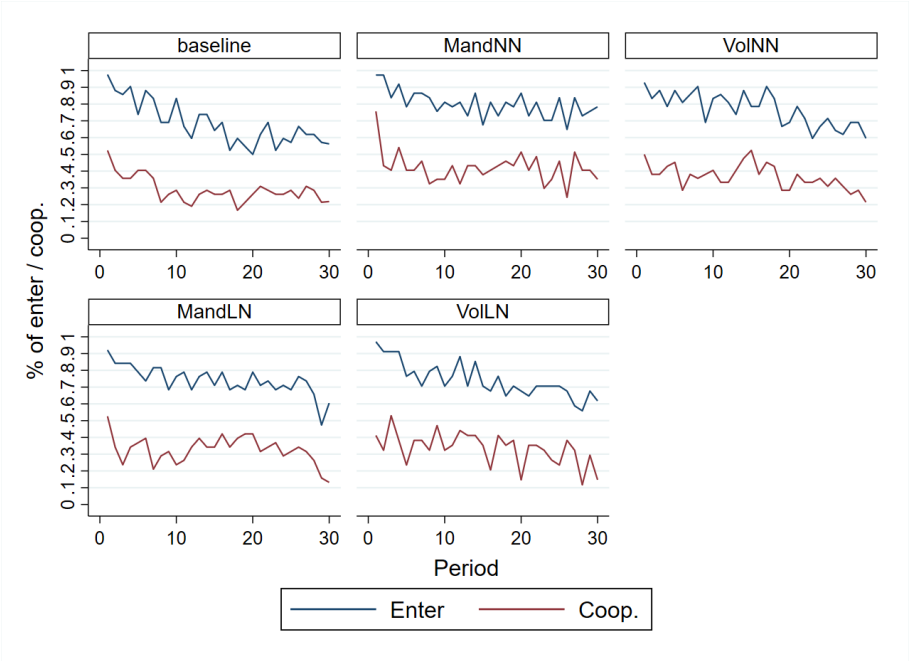


Figure D.3: The dynamic of P1's choices (period averages)

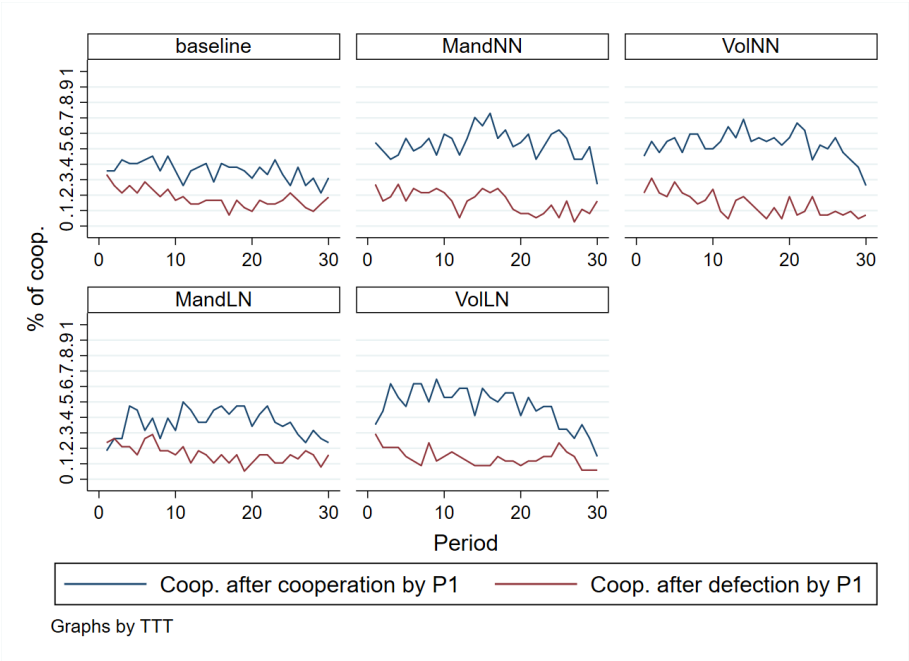


Figure D.4: The dynamic of P2's cooperation choices

Figure D.5: The dynamic of disclosure (rate averaged at period level).

D.3 Analysis of the 4 possibles strategies of P2

Figure D.6 reports the distribution of pure strategies of P2s across treatments. In Table D.1, we report the marginal effect of a multinomial logit regression explaining the strategy choice of P2s by the factorial interaction of "noisy" and "Voluntary". The results suggest that, while "Voluntary" has no effect on strategy choice at all, noise reduces the likelihood of conditional cooperation by 10 pp and increases the likelihood of defection by the same amount.

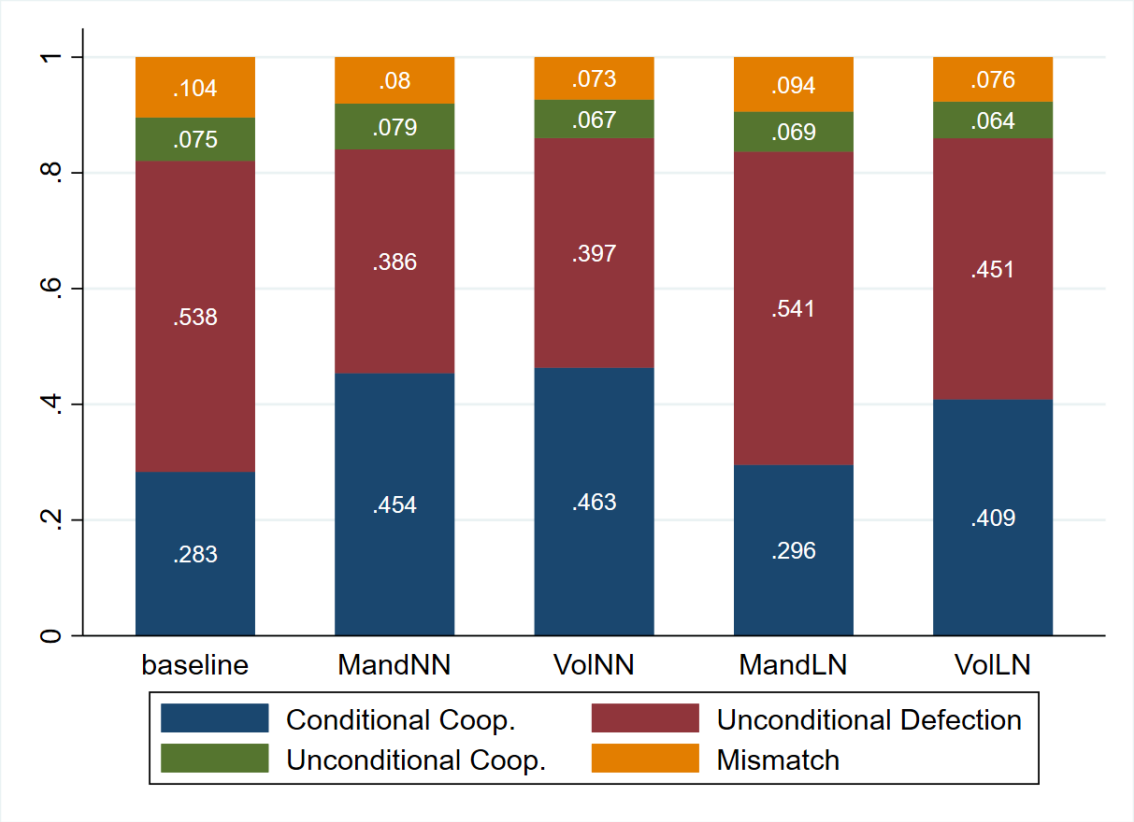


Figure D.6: The distribution of the 4 pure strategies of P2, by Treatment

	(1)
Marginal effect of "Voluntary"	
Conditional Cooperation	0.038 (0.049)
Unconditional Cooperation	-0.005 (0.012)
Unconditional Defection	-0.019 (0.051)
Mismatch	-0.014 (0.025)
Marginal effect of "noisy"	
Conditional Cooperation	-0.106*** (0.039)
Unconditional Cooperation	-0.010 (0.013)
Unconditional Defection	0.107*** (0.040)
Mismatch	0.009 (0.027)
Observations	5780

Standard errors in parentheses are clustered at the session level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table D.1: The treatment effect on the choice of pure strategy by P2 (multinomial logit model).

D.4 Individuals and behavioral types.

Does it pay to be skeptical? Our results suggest some skepticism among P1s. Here, we check whether being skeptical is beneficial in terms of earnings. To do so, we compute for each individual P1 the extent to which her cooperation rate depends on whether she received the information. This dependence of cooperation on information serves as measure of skepticism. We regress P1s' average stage-game earnings on this measure of skepticism, and we find that more skeptical individuals earn more (OLS with standard errors clustered at the session level: $b = 1.225,\ p = 0.03$).

Does it pay to reveal? P2 chose to reveal in the majority of the cases where he actually had a choice. We saw that disclosure is significantly linked to the quality of the record. Now we want to investigate individuals strategies. To do so, we

compute for each P2 in the Endo treatments the percentage of time he cooperates and the percentage of time he discloses. Using the kmeans algorithm, we identify 2 clusters with a natural interpretation: the first cluster is composed of individuals who barely cooperate and disclose (n=33). The second cluster consists of individuals who cooperate and disclose most of the time (n=43).[7] Both these strategies make sense: if a P2 believes that P1s are not very skeptical, the first strategy makes sense (Jin et al. (2021) find evidence of such beliefs in a sender-receiver game). Conversely, the second strategy makes sense for a P2 who anticipates that the P1s are skeptical. Figure D.7 represent the 2 clusters. We checked which strategy is the most beneficial for a P2, and we found that participants in the second cluster had higher average stage-game profits (7.42 vs 6.48, $p = 0.0017$ in a Mann-Whitney test).

[7]The same message holds if we choose to identify 3 clusters.

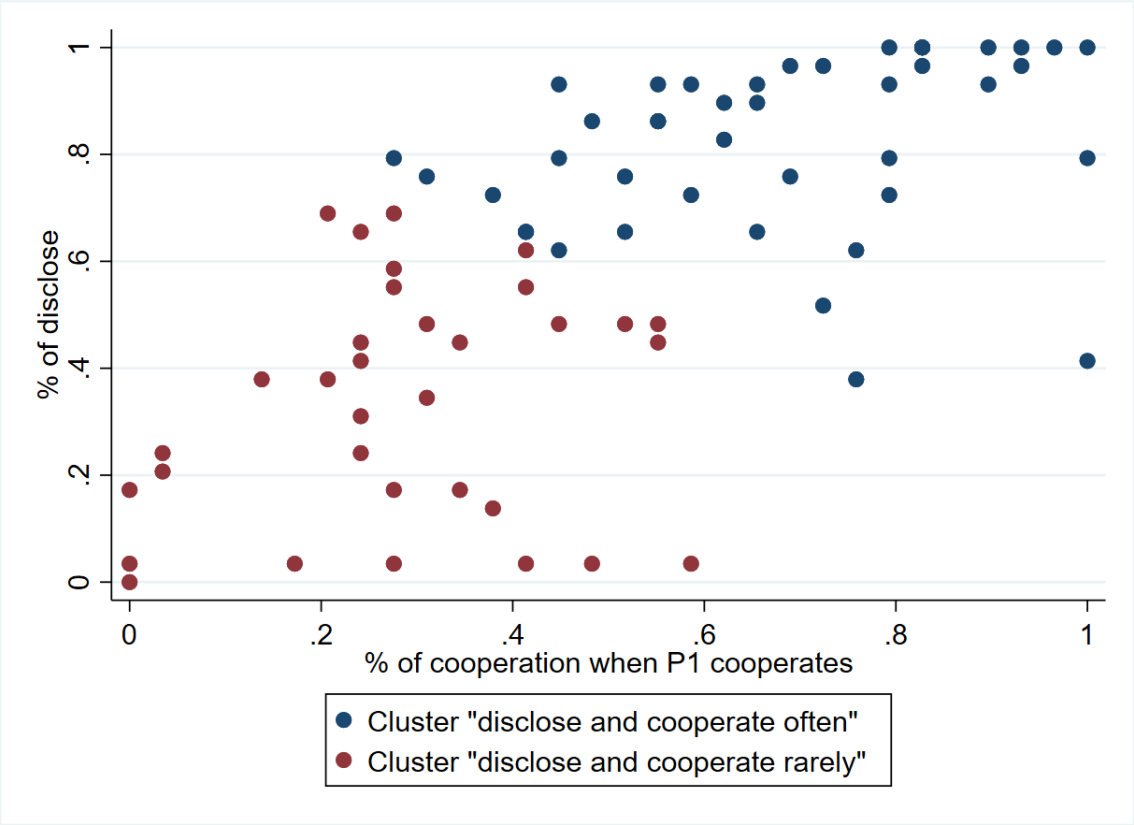


Figure D.7: Cluster of P2s according to disclosure and cooperation.

E (Seemingly) irrational behavior.

Some behavior are hard to rationalize at first sight: In a lot of cases, P1s enter but do not cooperate (overentry) and P2s cooperate even if P1 defected. In this section, we discuss whether such behavior is indeed an indication for irrationality.

Overentry by P1. First, "over"entry by P1 might actually be motivated by social curiosity (one learns the corresponding conditional choice of P2 when entering). In addition, P1s might believe that (there is a small probability that) P2s cooperate in response to defection. Even if this probability is small, this can explain the aforementioned choices: Let's say that P1 believes that P2s cooperate when P1

cooperates with probability q. Let's say that P1 believes that P2s cooperate when P1 defects with probability p. Given our parameters, two conditions must be met so that P1 enters AND defects:

(1): 20p-3(1-p) ≥ 5

(2) : $20p + 3(1-p) \geq +(1-q)$

(1) is the condition to "entering conditional on defecting" It imposes $p > \frac{2}{17} \sim 0.11$. (2) is the condition "defecting conditional on entering". It imposes $p > (9q-2)/17$ There is a large set of (p, q) satisfying (1) & (2). Some are à-priori unlikely ($p > q$), others are more plausible (e.g. $p \geq .15$ & $q \leq .3$ or $p \geq .2$ & $q \leq .5$). Miettinen et al. (2020) for instance elicit the beliefs of participants in a sequential prisoner dilemma and find that first movers on average expect second movers to cooperate 50% of the time when P1 cooperates, and 20% of the time when P1 defects. This might be due to P1s misunderstanding the game, to P1s expecting errors by P2s or to P1s expecting that P2 might be concerned by social welfare.

Cooperation after defection. "cooperate when P1 defected" can be due to unconditional cooperation, i.e. P2s who cooperate irrespective of the decision of P1, motivated by e.g. altruism or social welfare. By choosing to cooperate to a defector, one increases the social welfare 3.5 folds (from 6 to 21) or by 15 ECUs (75cents). Such concern for welfare is well documented in the literature (see e.g. Engelmann and Strobel, 2004). Miettinen et al. (2020) find that "concern for social welfare" is a good explanation (along with reciprocity, among others) of behaviors in the sequential prisoner dilemma. Interestingly enough, they also find that more or less 15% of

P2s "cooperate when P1 defects". This should ease our concerns about irrationality on P2s' behalf.

On the other hand, mismatch, i.e. P2 who choose to cooperate only when P1 defected is more puzzling. In this situation, reciprocity, social welfare concerns or any other theory are of little help. Overall, "mismatch" choices correspond to 8,5% of P2s decisions. This type of behavior is concentrated on a small number of participants: 50% of P2s never mismatch. 75% of P2s mismatch less 5 times. Note that there is a significant negative time trend in mismatching, which is good news (people learn). Our data is overall comparable to the data in Miettinen et al. (2020). Our results are robust to excluding mismatch decisions, or to the exclusion of the 25% of P2 who mismatched 5 times or more.

F Instructions

[Baseline]

Thank you for participating in this experiment on decision making. You are not allowed to communicate with other participants during the entire session. Please turn off your cell phone. If you have a question, please raise your hand and wait until an experimenter comes to you to answer your question in private.

For showing up on time, you receive a €3 show-up fee. In addition, your decisions during the experiment earn you money. During the session, earnings will be expressed in terms of Experimental Currency Units, ECUs. ECUs convert to euro at the following rate: 20 ECUs = €1. You will be paid by bank transfer shortly after the experiment. Your decisions in the experiment will remain anonymous.

****

Description of the decisions.

At the beginning of the session, each participant is assigned the role of PLAYER 1 or PLAYER 2, once and for all.

The experiment consists of 30 rounds. For each round, the computer program randomly forms pairs composed of one PLAYER 1 and one PLAYER 2, who will have the possibility to participate in an interaction.

In this interaction, PLAYER 1 and PLAYER 2 will have simple decisions to make: choosing either “UP” or “DOWN” according to a procedure described hereafter. These decisions will impact both their own earnings as well as the other PLAYER’s earnings as summarized in Table 1.

Table 1 : Earning consequences of Player 1 and Player 2's decisions in the interaction.

	PLAYER 2 chooses “DOWN”	PLAYER 2 choses “UP”
PLAYER 1 chooses ‘’DOWN’’	• Player 1 earns 3 ECUs • Player 2 earns 3 ECUs	• Player 1 earns 20 ECUs • Player 2 earns 1 ECUs
PLAYER 1 chooses ‘’UP’’	• Player 1 earns 1 ECUs • Player 2 earns 20 ECUs	• Player 1 earns 10 ECUs • Player 2 earns 10 ECUs

However, the interaction might not take place in every round: at the beginning of each round, PLAYER 1 has to decide whether to participate or not, using the interface shown in Screenshot 1.

If PLAYER 1 decides to participate, he then chooses between “UP” and “DOWN”.

If PLAYER 1 decides not to participate, he has no further decision to make in this round, and the decisions of PLAYER 2 have no impact on earnings: PLAYER 1 and PLAYER 2 both earn 5 ECUs.

PLAYER 2 is not informed of the decisions of PLAYER 1 at this point and as a consequence makes two decisions. One decision is made ‘’as if” PLAYER 1 had decided to participate, and had subsequently chosen “UP”. The other decision is made “as if” PLAYER 1 had decided to participate, and had subsequently chosen “DOWN”. PLAYER 2 makes these decisions using the interface shown in Screenshot 2.

To compute earnings, only the decision of PLAYER 2 corresponding to the actual decision of PLAYER 1 is used. In other words:

- If PLAYER 1 chose not to participate, none of the decisions of PLAYER 2 impact earnings and both PLAYERS earn 5 ECUs.
- If PLAYER 1 chose to participate and “DOWN”, only the decision of PLAYER 2 made “as if” PLAYER1 had chosen “DOWN” matters. In this case, the consequence of PLAYER 2’s decision on earnings is found in the first line of Table 1.
- If PLAYER 1 chose to participate and “UP”, only the decision of PLAYER 2 made “as if” PLAYER1 had chosen “UP” matters. In this case, the consequence of PLAYER 2’s decision on earnings is found in the second line of Table 1.

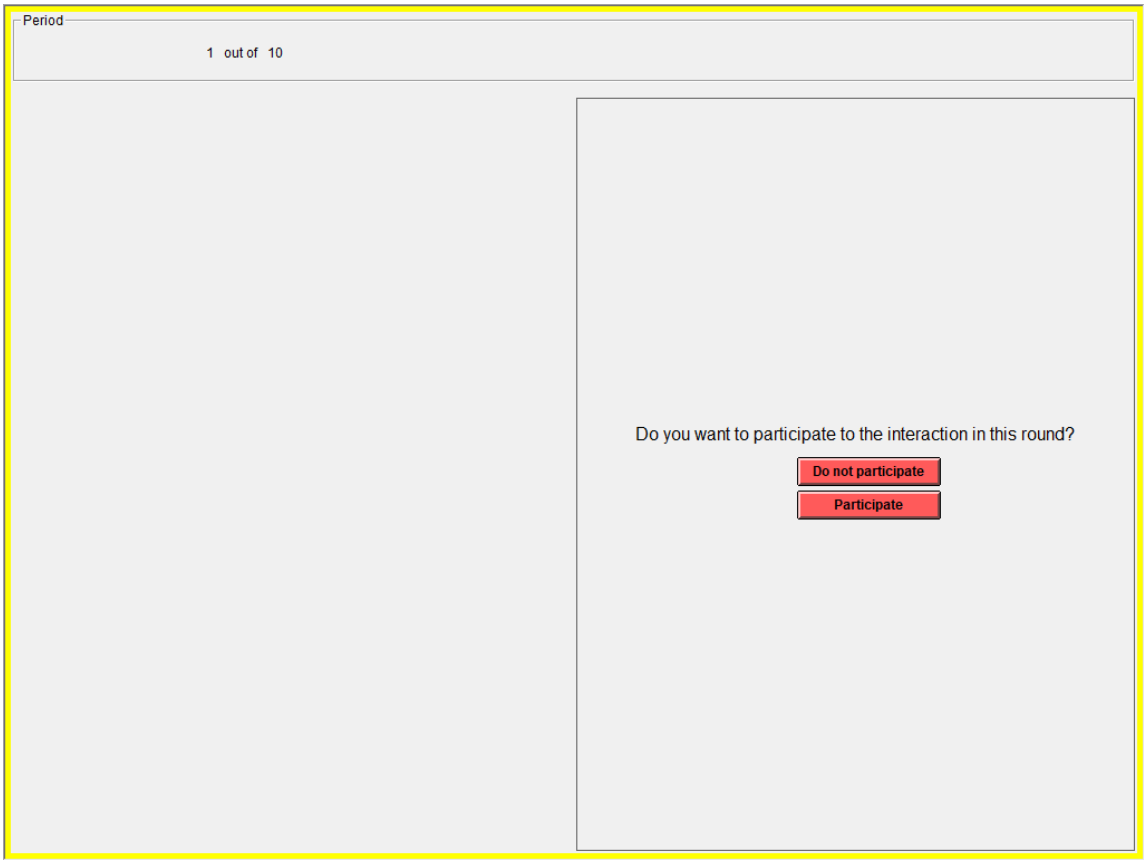
At the end of each round, you are informed of your earnings for the present round and of the earnings you have accumulated up to this round. You are not informed of the decision of the other PLAYER.

****

Before the first round of the experiment, you will be asked to answer several questions aimed at checking your understanding of the decisions you will have to make.

After the last round, you will have to answer a demographic questionnaire. This questionnaire does not threaten the anonymity of your decisions.

Please read these instructions again before clicking OK to proceed to the experiment. Should you have any question, silently raise your hand and an experimenter will come to answer privately.



Screenshot 1: Participation decision (PLAYER 1)

Screenshot 2: UP or DOWN decisions (PLAYER 2)

[Mandatory No Noise]

Thank you for participating in this experiment on decision making. You are not allowed to communicate with other participants during the entire session. Please turn off your cell phone. If you have a question, please raise your hand and wait until an experimenter comes to you to answer your question in private.

For showing up on time, you receive a €3 show-up fee. In addition, your decisions during the experiment earn you money. During the session, earnings will be expressed in terms of Experimental Currency Units, ECU. ECU converts to euro at the following rate: 20 ECUs = €1. You will be paid by bank transfer shortly after the experiment. Your decisions in the experiment will remain anonymous.

****

Description of the decisions.

At the beginning of the session, each participant is assigned to the role of PLAYER 1 or PLAYER 2, once and for all.

The experiment consists of 30 rounds. For each round, the computer program randomly forms pairs composed of one PLAYER 1 and one PLAYER 2, who will have the possibility to participate in an interaction.

In this interaction, PLAYER 1 and PLAYER 2 will have simple decisions to make: choosing either "UP" or "DOWN" according to a procedure described hereafter. These decisions will impact both their own earnings as well as the other PLAYER's earnings as summarized in Table 1.

Table 1: Earning consequences of Player 1 and Player 2's decisions in the interaction.

	PLAYER 2 chooses "DOWN"	PLAYER 2 choses "UP"
PLAYER 1 chooses ''down''	• Player 1 earns 3 ECUs • Player 2 earns 3 ECUs	• Player 1 earns 20 ECUs • Player 2 earns 1 ECUs
PLAYER 1 chooses ''up''	• Player 1 earns 1 ECUs • Player 2 earns 20 ECUs	• Player 1 earns 10 ECUs • Player 2 earns 10 ECUs

However, the interaction might not take place in every round: at the beginning of each round, PLAYER 1 has to decide whether to participate or not, using the interface shown in Screenshot 1.

To help PLAYER 1's participation decision, from round 2 on, the past decisions of PLAYER 2 will be disclosed on the left panel of PLAYER 1's decision screen, as shown in screenshot 2.

If PLAYER 1 decides to participate, he then chooses between "UP" and "DOWN".

If PLAYER 1 decides not to participate, he has no further decision to make in this round, and the decisions of PLAYER 2 have no impact on earnings: PLAYER 1 and PLAYER 2 both earn 5 ECUs.

PLAYER 2 is not informed of the decisions of PLAYER 1 at this point and as a consequence makes two decisions. One decision is made ''as if" PLAYER 1 had decided to participate, and had subsequently chosen "UP". The other decision is made "as if" PLAYER 1 had decided to participate, and had subsequently chosen "DOWN". PLAYER 2 makes these decisions using the interface shown in Screenshot 3.

To compute earnings, only the decision of PLAYER 2 corresponding to the actual decisions of PLAYER 1 is used. In other words:

- If PLAYER 1 chose not to participate, none of the decisions of PLAYER 2 impact earnings and both PLAYERS earn 5 ECUs.
- If PLAYER 1 chose to participate and “DOWN”, only the decision of PLAYER 2 taken “as if” PLAYER1 had chosen “DOWN” matters. In this case, the consequence of PLAYER 2’s decision on earnings is found in the first line of Table 1.
- If PLAYER 1 chose to participate and “UP”, only the decision of PLAYER 2 taken “as if” PLAYER1 had chosen “UP” matters. In this case, the consequence of PLAYER 2’s decision on earnings is found in the second line of Table 1.

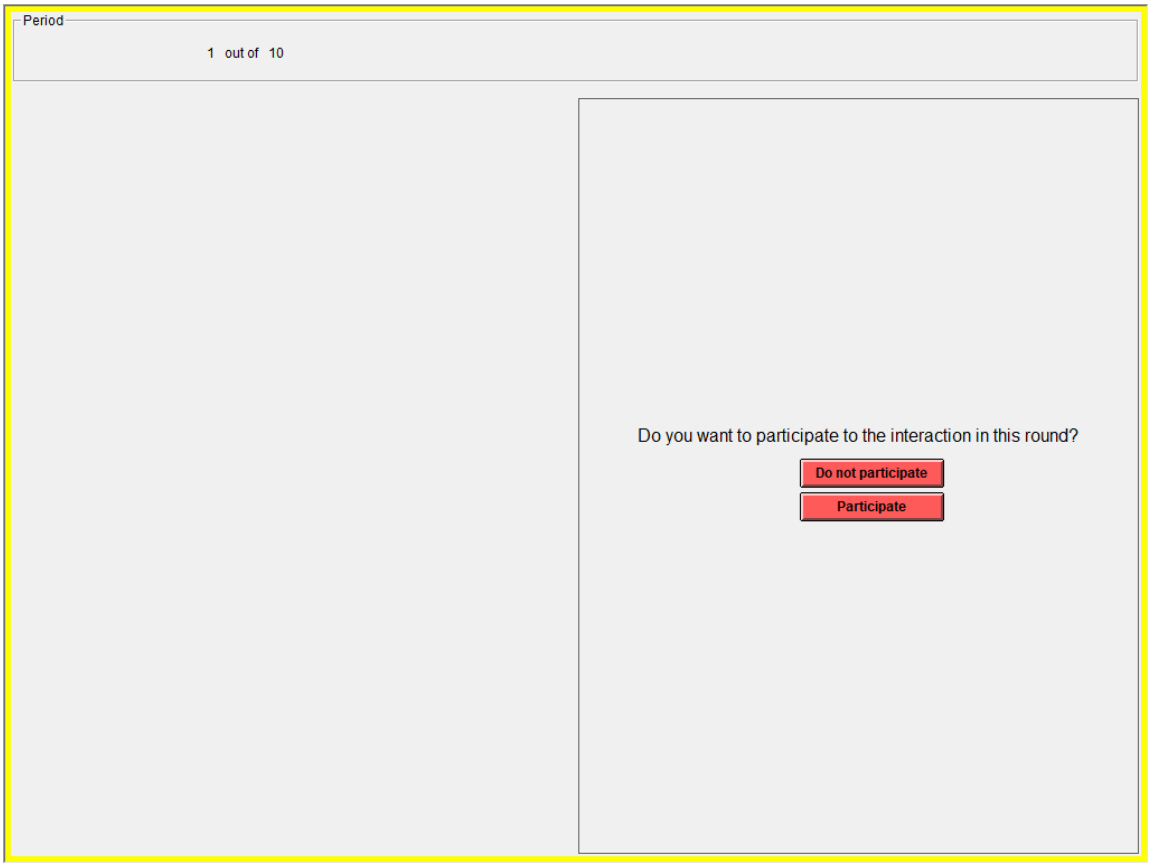
At the end of each round, you are informed of your earnings for the present round and of the earnings you have accumulated up to this round. You are not informed of the decision of the other PLAYER.

****

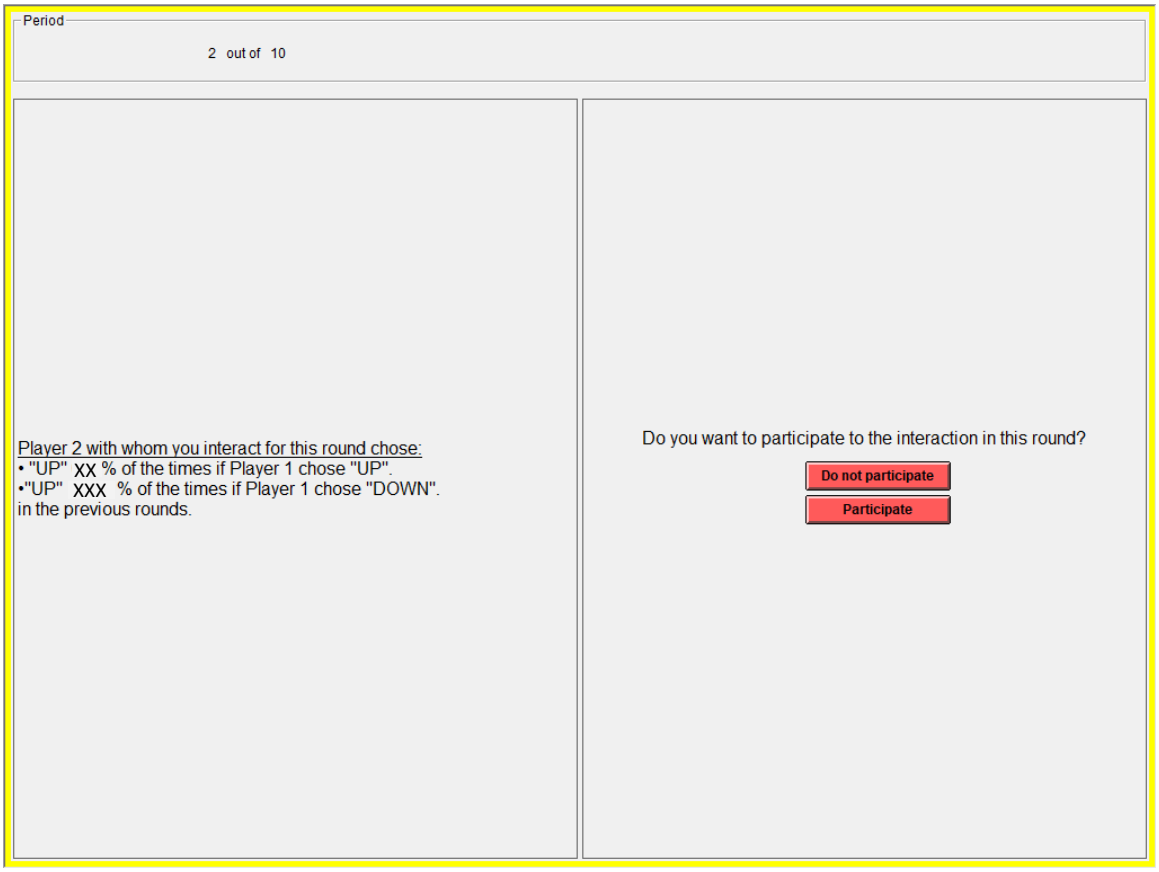
Before the first round of the experiment, you will be asked to answer several questions aimed at checking your understanding of the decisions you will have to make.

After the last round, you will have to answer a demographic questionnaire. This questionnaire does not threaten the anonymity of your decisions.

Please read these instructions again before clicking OK to proceed to the experiment. Should you have any question, silently raise your hand and an experimenter will come to answer privately.



Screenshot 1: Participation decision (PLAYER 1)



Screenshot 2: Participation decision with information disclosure (PLAYER 1)

Screenshot 3: UP or DOWN decisions (PLAYER 2)

[Mandatory Noise]

Thank you for participating in this experiment on decision making. You are not allowed to communicate with other participants during the entire session. Please turn off your cell phone. If you have a question, please raise your hand and wait until an experimenter comes to you to answer your question in private.

For showing up on time, you receive a €3 show-up fee. In addition, your decisions during the experiment earn you money. During the session, earnings will be expressed in terms of Experimental Currency Units, ECU. ECU converts to euro at the following rate: 20 ECUs = €1. You will be paid by bank transfer shortly after the experiment. Your decisions in the experiment will remain anonymous.

****

Description of the decisions.

At the beginning of the session, each participant is assigned to the role of PLAYER 1 or PLAYER 2, once and for all.

The experiment consists of 30 rounds. For each round, the computer program randomly forms pairs composed of one PLAYER 1 and one PLAYER 2, who will have the possibility to participate in an interaction.

In this interaction, PLAYER 1 and PLAYER 2 will have simple decisions to make: choosing either "UP" or "DOWN" according to a procedure described hereafter. These decisions will impact both their own earnings as well as the other PLAYER's earnings as summarized in Table 1.

Table 1: Earning consequences of Player 1 and Player 2's decisions in the interaction.

	PLAYER 2 chooses "DOWN"	PLAYER 2 choses "UP"
PLAYER 1 chooses ''down''	• Player 1 earns 3 ECUs • Player 2 earns 3 ECUs	• Player 1 earns 20 ECUs • Player 2 earns 1 ECUs
PLAYER 1 chooses ''up''	• Player 1 earns 1 ECUs • Player 2 earns 20 ECUs	• Player 1 earns 10 ECUs • Player 2 earns 10 ECUs

However, the interaction might not take place in every round: at the beginning of each round, PLAYER 1 has to decide whether to participate or not, using the interface shown in Screenshot 1.

To help PLAYER 1's participation decision, from round 2 on, there are 9 chances out of 10 that the past decisions of PLAYER 2 get disclosed on the left panel of PLAYER 1's decision screen, as shown in screenshot 2. There is 1 chance out of 10 than nothing gets disclosed, as in Screenshot 1.

If PLAYER 1 decides to participate, he then chooses between "UP" and "DOWN".

If PLAYER 1 decides not to participate, he has no further decision to make in this round, and the decisions of PLAYER 2 have no impact on earnings: PLAYER 1 and PLAYER 2 both earn 5 ECUs.

PLAYER 2 is not informed of the decisions of PLAYER 1 at this point and as a consequence makes two decisions. One decision is made ''as if" PLAYER 1 had decided to participate, and had subsequently chosen "UP". The other decision is made "as if" PLAYER 1 had decided to participate, and had subsequently chosen "DOWN". PLAYER 2 makes these decisions using the interface shown in Screenshot 3.

To compute earnings, only the decision of PLAYER 2 corresponding to the actual decisions of PLAYER 1 is used. In other words:

- If PLAYER 1 chose not to participate, none of the decisions of PLAYER 2 impact earnings and both PLAYERS earn 5 ECUs.
- If PLAYER 1 chose to participate and "DOWN", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "DOWN" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the first line of Table 1.
- If PLAYER 1 chose to participate and "UP", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "UP" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the second line of Table 1.

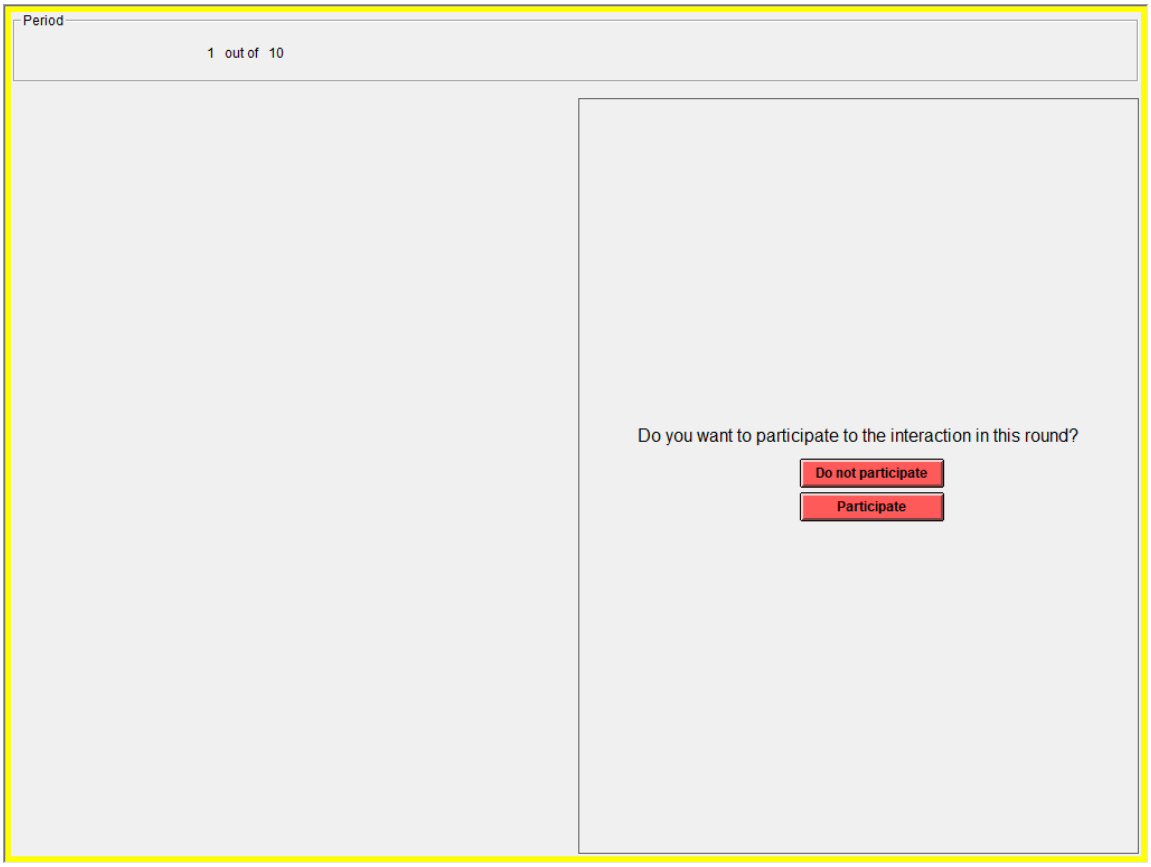
At the end of each round, you are informed of your earnings for the present round and of the earnings you have accumulated up to this round. You are not informed of the decision of the other PLAYER.

****

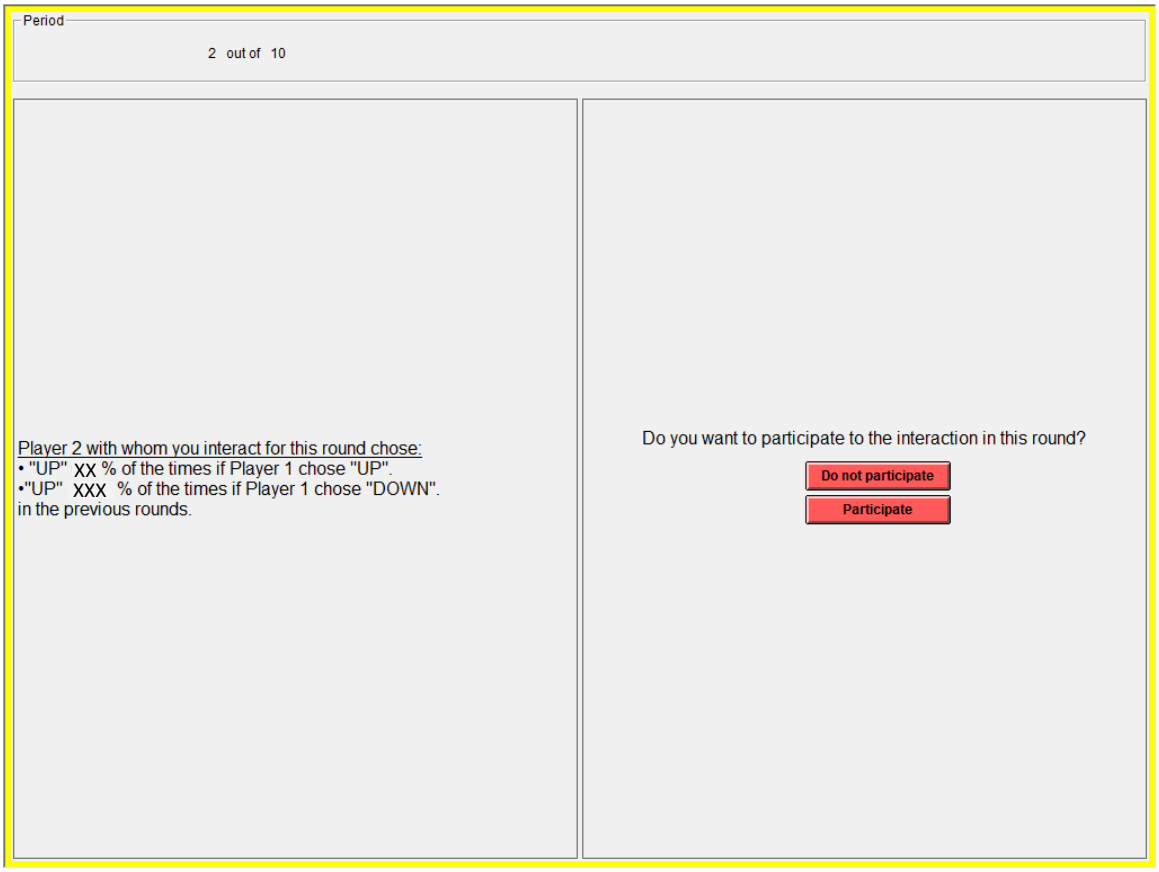
Before the first round of the experiment, you will be asked to answer several questions aimed at checking your understanding of the decisions you will have to make.

After the last round, you will have to answer a demographic questionnaire. This questionnaire does not threaten the anonymity of your decisions.

Please read these instructions again before clicking OK to proceed to the experiment. Should you have any question, silently raise your hand and an experimenter will come to answer privately.



Screenshot 1: Participation decision (PLAYER 1)



Screenshot 2: Participation decision with information disclosure (PLAYER 1)

Screenshot 3: UP or DOWN decisions (PLAYER 2)

[Voluntary No Noise]

Thank you for participating in this experiment on decision making. You are not allowed to communicate with other participants during the entire session. Please turn off your cell phone. If you have a question, please raise your hand and wait until an experimenter comes to you to answer your question in private.

For showing up on time, you receive a €3 show-up fee. In addition, your decisions during the experiment earn you money. During the session, earnings will be expressed in terms of Experimental Currency Units, ECU. ECU converts to euro at the following rate: 20 ECUs = €1. You will be paid by bank transfer shortly after the experiment. Your decisions in the experiment will remain anonymous.

****

Description of the decisions.

At the beginning of the session, each participant is assigned to the role of PLAYER 1 or PLAYER 2, once and for all.

The experiment consists of 30 rounds. For each round, the computer program randomly forms pairs composed of one PLAYER 1 and one PLAYER 2, who will have the possibility to participate in an interaction.

In this interaction, PLAYER 1 and PLAYER 2 will have simple decisions to make: choosing either "UP" or "DOWN" according to a procedure described hereafter. These decisions will impact both their own earnings as well as the other PLAYER's earnings as summarized in Table 1.

Table 1 : Earning consequences of Player 1 and Player 2's decisions in the interaction.

	PLAYER 2 chooses "DOWN"	PLAYER 2 choses "UP"
PLAYER 1 chooses ''down''	• Player 1 earns 3 ECUs • Player 2 earns 3 ECUs	• Player 1 earns 20 ECUs • Player 2 earns 1 ECUs
PLAYER 1 chooses ''up''	• Player 1 earns 1 ECUs • Player 2 earns 20 ECUs	• Player 1 earns 10 ECUs • Player 2 earns 10 ECUs

However, the interaction might not take place in every round: at the beginning of each round, PLAYER 1 has to decide whether to participate or not, using the interface shown in Screenshot 1.

To help PLAYER 1's participation decision, from round 2 on PLAYER 2 can choose to disclose his past decisions (see Screenshot 2). If so, the past decisions of PLAYER 2 will be disclosed on the left part of PLAYER 1's decision screen, as shown in Screenshot 3.

If PLAYER 1 decides to participate, he then chooses between "UP" and "DOWN".

If PLAYER 1 decides not to participate, he has no further decision to make in this round, and the decisions of PLAYER 2 have no impact on earnings: PLAYER 1 and PLAYER 2 both earn 5 ECUs.

PLAYER 2 is not informed of the decisions of PLAYER 1 at this point and as a consequence makes two decisions. One decision is made ''as if" PLAYER 1 had decided to participate, and had subsequently chosen "UP". The other decision is made "as if" PLAYER 1 had decided to participate, and had subsequently chosen "DOWN". PLAYER 2 makes these decisions using the interface shown in Screenshot 4.

To compute earnings, only the decision of PLAYER 2 corresponding to the actual decisions of PLAYER 1 is used. In other words:

- If PLAYER 1 chose not to participate, none of the decision of PLAYER 2 impacts earnings and both PLAYERS earn 5 ECUs.
- If PLAYER 1 chose to participate and "DOWN", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "DOWN" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the first line of Table 1.
- If PLAYER 1 chose to participate and "UP", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "UP" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the second line of Table 1.

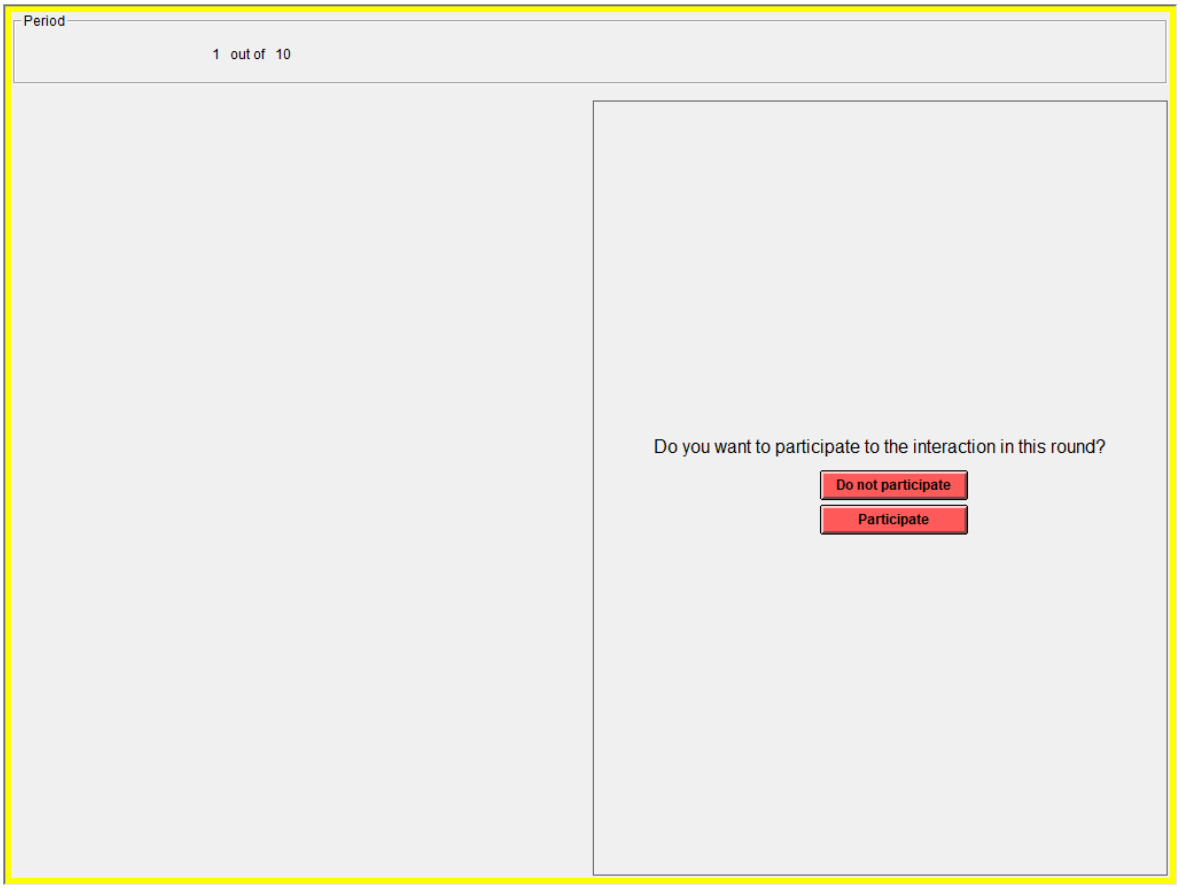
At the end of each round, you are informed of your earnings for the present round and of the earnings you have accumulated up to this round. You are not informed of the decision of the other PLAYER.

****

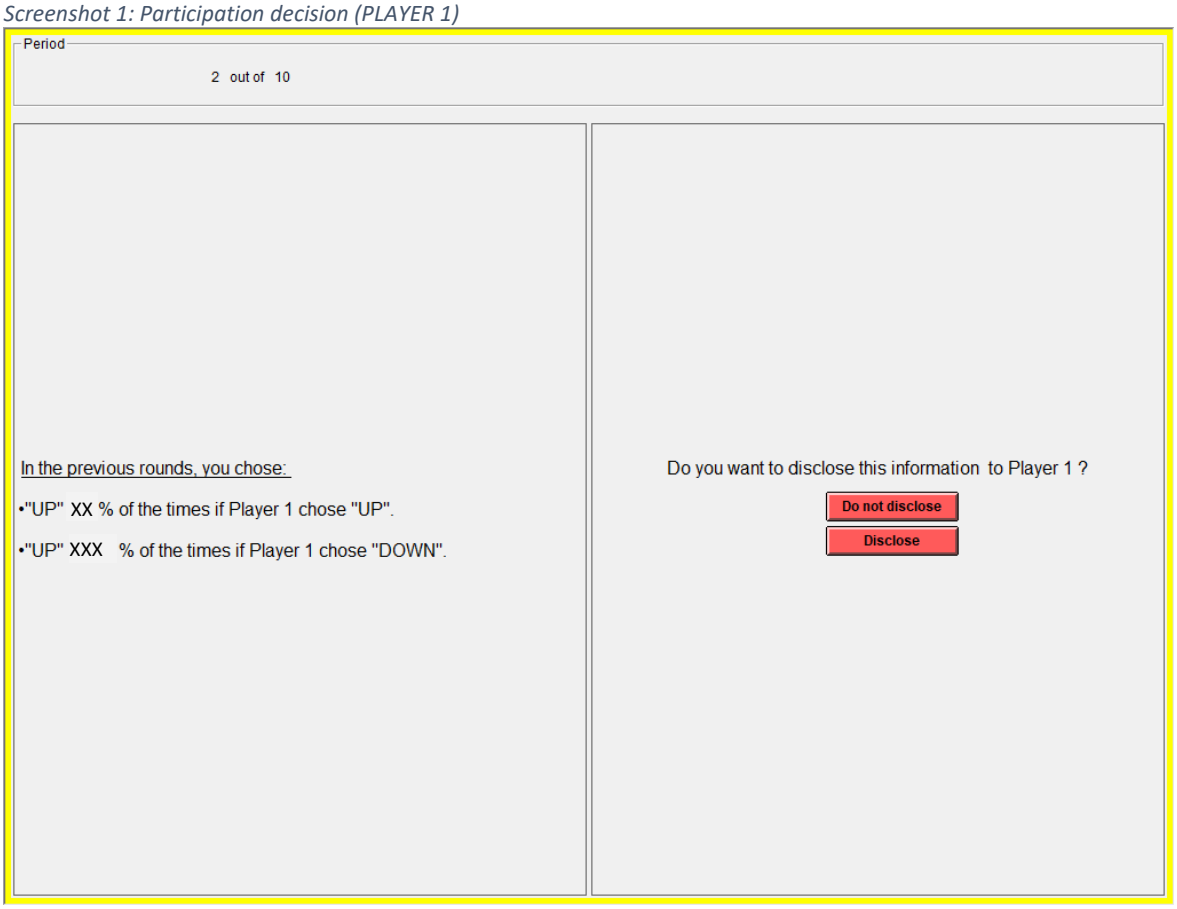
Before the first round of the experiment, you will be asked to answer several questions aimed at checking your understanding of the decisions you will have to make.

After the last round, you will have to answer a demographic questionnaire. This questionnaire does not threaten the anonymity of your decisions.

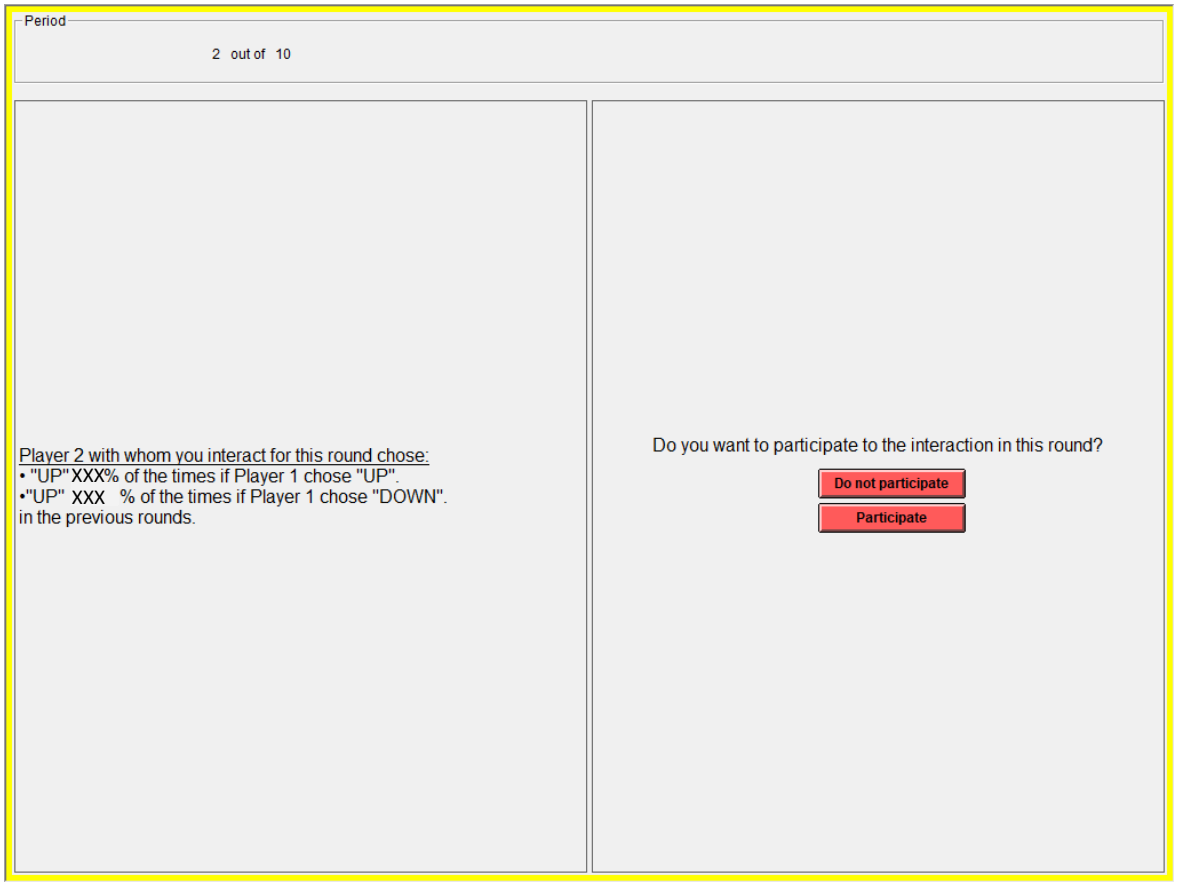
Please read these instructions again before clicking OK to proceed to the experiment. Should you have any question, silently raise your hand and an experimenter will come to answer privately.



Screenshot 1: Participation decision (PLAYER 1)



Screenshot 2: Disclosure decision (PLAYER 2)



Screenshot 3: Participation decision with information disclosure (PLAYER 1)

Screenshot 4: UP or DOWN decisions (PLAYER 2)

[Voluntary Noise]

Thank you for participating in this experiment on decision making. You are not allowed to communicate with other participants during the entire session. Please turn off your cell phone. If you have a question, please raise your hand and wait until an experimenter comes to you to answer your question in private.

For showing up on time, you receive a €3 show-up fee. In addition, your decisions during the experiment earn you money. During the session, earnings will be expressed in terms of Experimental Currency Units, ECU. ECU converts to euro at the following rate: 20 ECUs = €1. You will be paid by bank transfer shortly after the experiment. Your decisions in the experiment will remain anonymous.

****

Description of the decisions.

At the beginning of the session, each participant is assigned to the role of PLAYER 1 or PLAYER 2, once and for all.

The experiment consists of 30 rounds. For each round, the computer program randomly forms pairs composed of one PLAYER 1 and one PLAYER 2, who will have the possibility to participate in an interaction.

In this interaction, PLAYER 1 and PLAYER 2 will have simple decisions to make: choosing either "UP" or "DOWN" according to a procedure described hereafter. These decisions will impact both their own earnings as well as the other PLAYER's earnings as summarized in Table 1.

Table 1 : Earning consequences of Player 1 and Player 2's decisions in the interaction.

	PLAYER 2 chooses "DOWN"	PLAYER 2 choses "UP"
PLAYER 1 chooses ''down''	• Player 1 earns 3 ECUs • Player 2 earns 3 ECUs	• Player 1 earns 20 ECUs • Player 2 earns 1 ECUs
PLAYER 1 chooses ''up''	• Player 1 earns 1 ECUs • Player 2 earns 20 ECUs	• Player 1 earns 10 ECUs • Player 2 earns 10 ECUs

However, the interaction might not take place in every round: at the beginning of each round, PLAYER 1 has to decide whether to participate or not, using the interface shown in Screenshot 1.

To help PLAYER 1's participation decision, from round 2 on PLAYER 2 can choose to disclose his past decisions (see Screenshot 2). If Player 2 chooses to disclose, there is 9 chances out of 10, that the past decisions of PLAYER 2 get disclosed on the left part of PLAYER 1's decision screen, as shown in Screenshot 3, and 1 chance out of 10 that nothing gets disclosed.

If PLAYER 1 decides to participate, he then chooses between "UP" and "DOWN".

If PLAYER 1 decides not to participate, he has no further decision to make in this round, and the decisions of PLAYER 2 have no impact on earnings: PLAYER 1 and PLAYER 2 both earn 5 ECUs.

PLAYER 2 is not informed of the decisions of PLAYER 1 at this point and as a consequence makes two decisions. One decision is made ''as if" PLAYER 1 had decided to participate, and had subsequently chosen "UP". The other decision is made "as if" PLAYER 1 had decided to participate, and had subsequently chosen "DOWN". PLAYER 2 makes these decisions using the interface shown in Screenshot 4.

To compute earnings, only the decision of PLAYER 2 corresponding to the actual decisions of PLAYER 1 is used. In other words:

- If PLAYER 1 chose not to participate, none of the decision of PLAYER 2 impacts earnings and both PLAYERS earn 5 ECUs.
- If PLAYER 1 chose to participate and "DOWN", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "DOWN" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the first line of Table 1.
- If PLAYER 1 chose to participate and "UP", only the decision of PLAYER 2 taken "as if" PLAYER1 had chosen "UP" matters. In this case, the consequence of PLAYER 2's decision on earnings is found in the second line of Table 1.

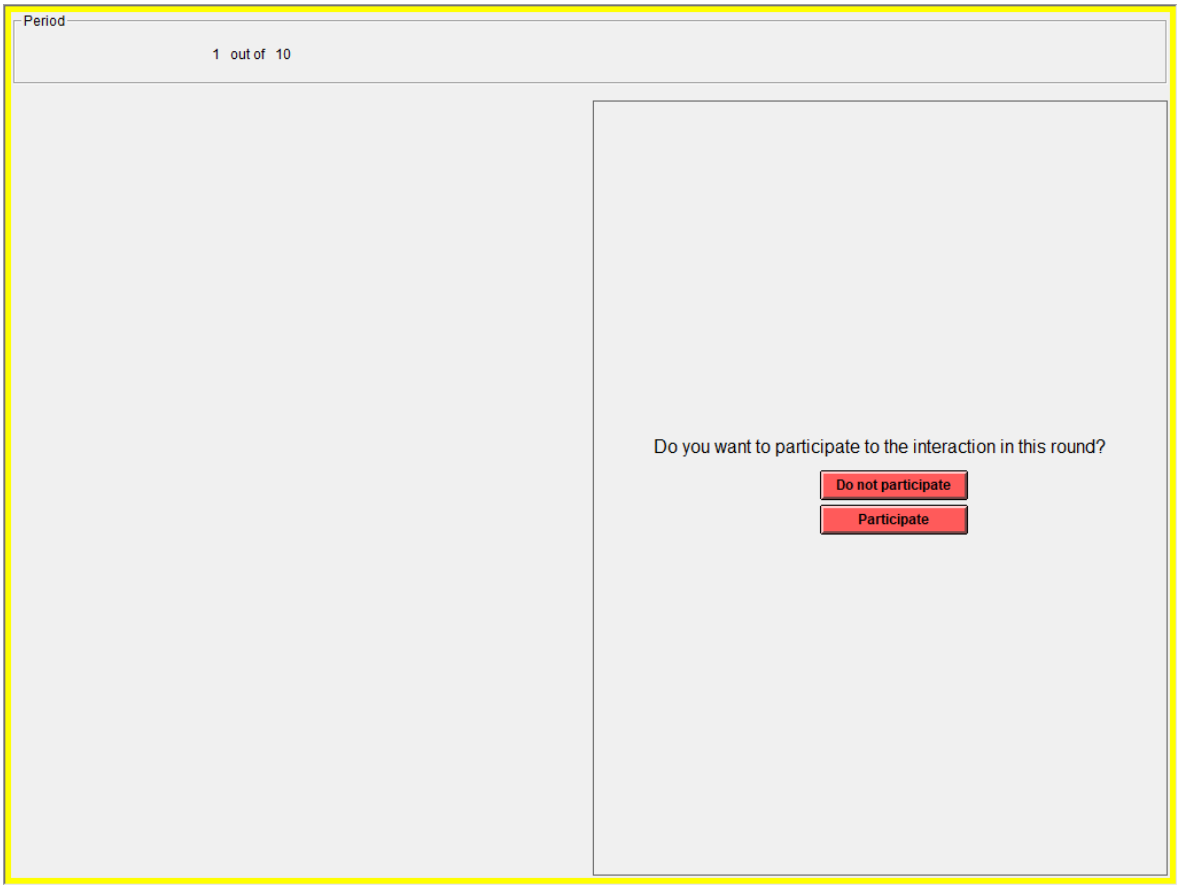
At the end of each round, you are informed of your earnings for the present round and your cumulated earnings, but not of the decision of the other PLAYER.

****

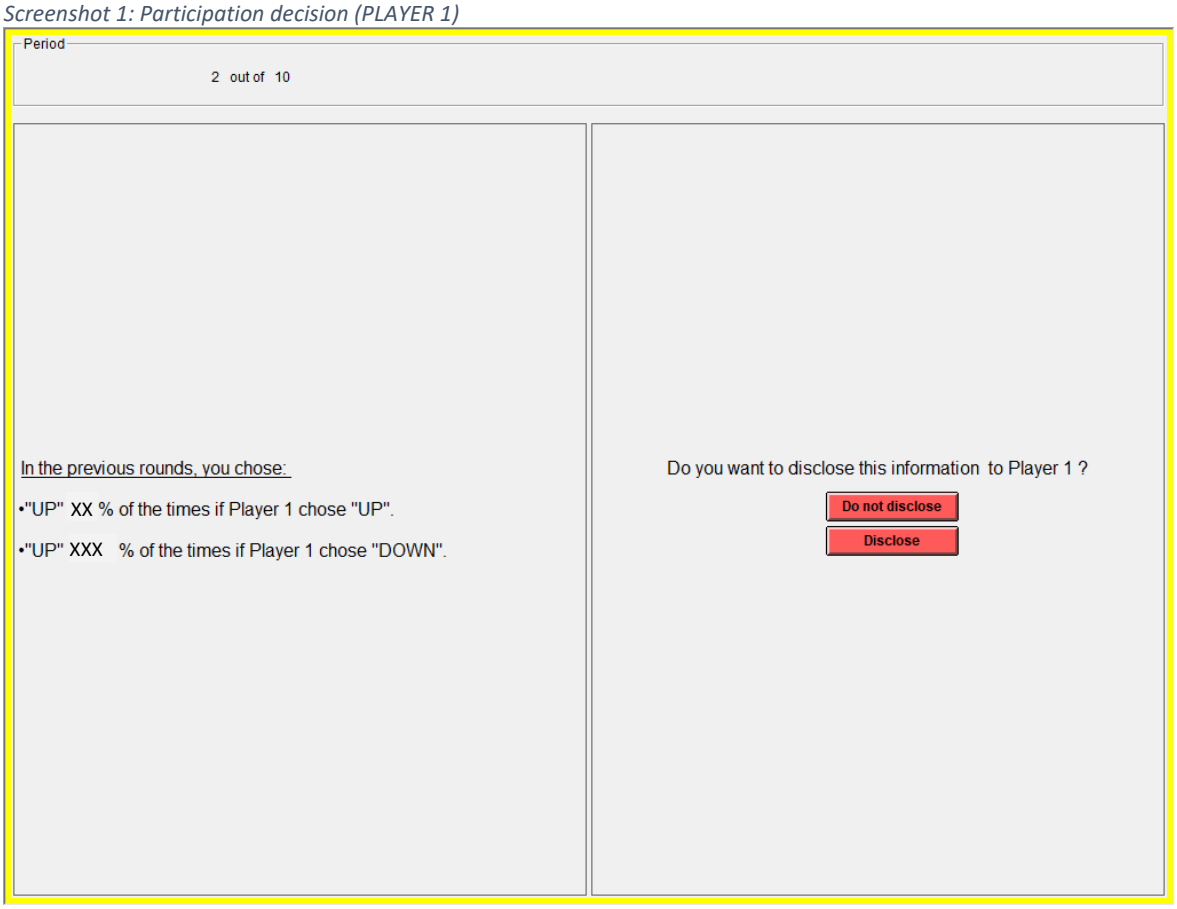
Before the first round of the experiment, you will be asked to answer several questions aimed at checking your understanding of the decisions you will have to make.

After the last round, you will have to answer a demographic questionnaire. This questionnaire does not threaten the anonymity of your decisions.

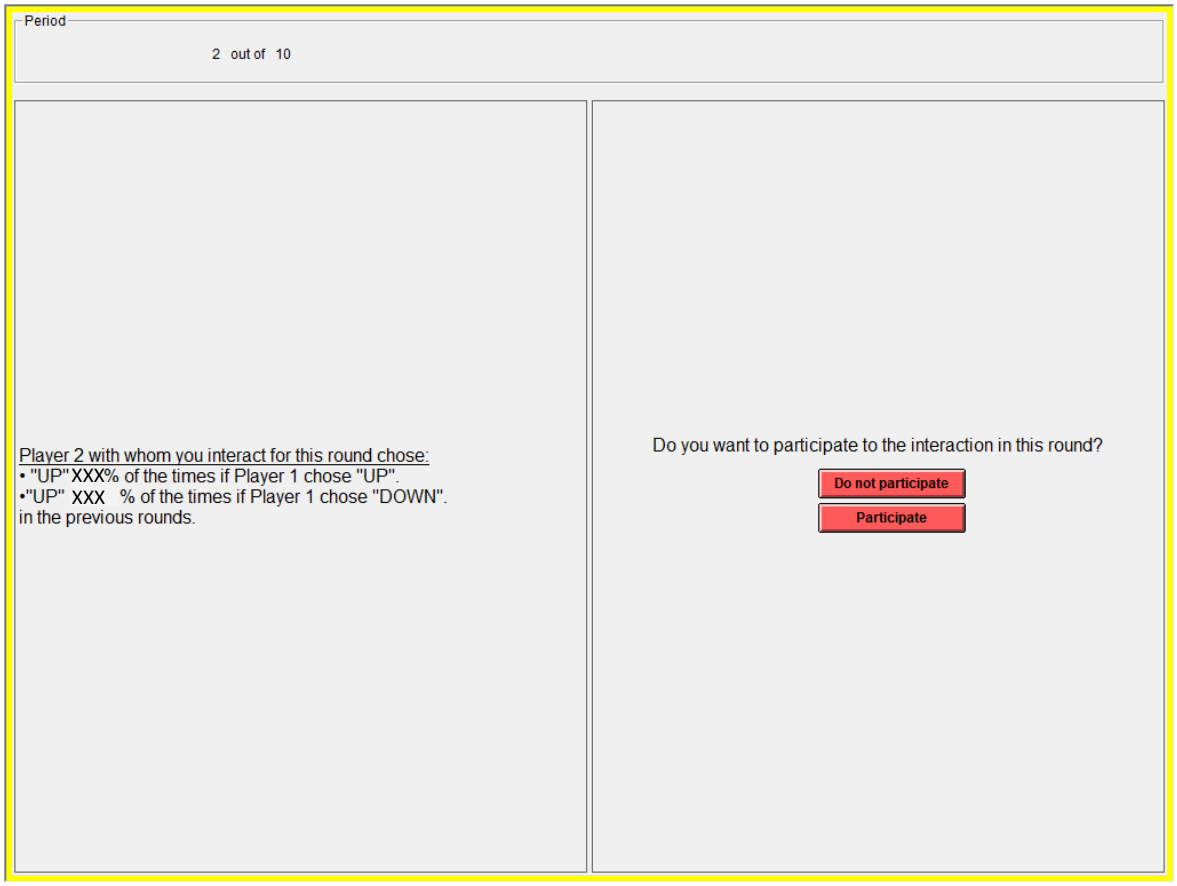
Please read these instructions again before clicking OK to proceed to the experiment. Should you have any question, silently raise your hand so that an experimenter will come to answer privately.



Screenshot 1: Participation decision (PLAYER 1)



Screenshot 2: Disclosure decision (PLAYER 2)



Screenshot 3: Participation decision with information disclosure (PLAYER 1)

Screenshot 4: UP or DOWN decisions (PLAYER 2)