Experimental Economics (2026), 1–18
doi:10.1017/eec.2026.10069

economic science association

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL PAPER

Sustaining cooperation with correlated information: an experimental test

Yongping Bao[1], Fabian Dvorak[2] and Sebastian Fehrler[1]

[1]Socium, University of Bremen: Universitat Bremen, Germany
[2]Department of Environmental Social Sciences, Eawag, Dübendorf, Switzerland
Corresponding author: Sebastian Fehrler; Email: sebastian.fehrler@uni-bremen.de

(Received 25 February 2025 UTC; revised 9 February 2026 UTC; accepted 29 September 2026 UTC)

Abstract

In infinitely or indefinitely repeated games with noisy signals about others' actions, sustaining cooperation is difficult. Theoretical work shows that cooperation can be maintained if the signals are correlated and the degree of correlation depends on the actions. In this study, we implement such an information structure in a laboratory experiment and investigate whether subjects are able to sustain cooperation by conditioning their behavior on it. A substantial number of subjects adopt strategies accounting for the correlation, but this does not increase cooperation compared to a control treatment without correlation, as behavior with independent signals is more lenient when players can communicate, while without communication cooperation remains low in both settings.

Keywords: cooperation; correlated signals; infinitely repeated games; monitoring; Prisoners' Dilemma; strategic uncertainty

JEL codes: C72; C73; C92; D83

1. Introduction

In a tightly controlled laboratory experiment, we study strategy choices and resulting cooperation rates in an indefinitely repeated Prisoners' Dilemma with imperfect public monitoring and correlated signals, where the degree of correlation depends on the actions of both players (as in Awaya & Krishna, 2016; 2019).

A large literature in game theory investigates cooperation in infinitely (or indefinitely) repeated games. In some applications, players monitor the actions of the opponents or partners perfectly. In many situations, however, opponents' actions are monitored only imperfectly. For example, in teamwork, players repeatedly put effort into a joint project. Their partners do not observe their actual effort, but instead the output which noisily represents the effort (Sekiguchi, 1997; Compte & Postlewaite, 2015). In the oligopoly model of Green & Porter (1984), firms observe the market price but not the output choices of their competitors. In both examples, shocks or other factors may confound the observed information, such that instead of directly observing the actual choice of the opponent, players only observe noisy signals that represent the opponents' actions. In these and many other settings with imperfect monitoring, payoffs naturally depend on observable outcomes rather than unobservable actions. For instance, in team production and professional partnerships, an individual's share of profits depends on measurable outputs of partners (e.g., billable hours, cases handled) rather than their unobservable effort. In supply chains, a buyer's payoff depends on the realized quality

© The Author(s), 2026. Published by Cambridge University Press on behalf of the Economic Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

of a delivered product, not on the supplier's effort in production. In principal-agent relationships, an employer's payoff depends on observable performance metrics rather than unobservable employee effort. In all these cases, what matters for payoffs is what agents observe or experience, not what others intended to do.

Understanding how cooperation can be sustained in indefinitely repeated games under imperfect monitoring has become an important topic of the literature, because sustaining cooperation is difficult when information is noisy. Awaya and Krishna (2016; 2019) point out that one way to sustain cooperation is through correlated information. While the literature commonly assumes that noisy signals solely depend on players' own chosen actions, they assume that the noisy signals depend on all actions. Under their assumption, the signals are correlated in a way that they are similar when players' actions are similar, and become dissimilar, as players' actions diverge. One example is a duopoly. The two firms have private information regarding their own sales, but their sales are correlated as they are sensitive to the prices on the market, and the degree of correlation depends on whether they chose similar actions.

The fundamental insight in Awaya and Krishna (2016; 2019) is that exploiting correlation in information can help to improve monitoring and sustain cooperation. In this study, we adopt and adapt their assumption of information correlation and explore the empirical implications. While their focus is on private monitoring, our design uses public monitoring.[1] They show that communication prior to every round can sustain cooperation based on pure strategies. A necessary condition is that the private signals are more strongly correlated if players choose the same action. By publicly reporting private signals, players can draw inferences about past actions based on these reports. Under public monitoring, signals are publicly observable and there is no role for truthful communication of signals. In this sense, our implementation represents the case of fully truthful communication under private monitoring.

Our central focus is thus not on the communication strategies players devise when they know how to exploit information contained in correlation. In contrast, we focus on the question whether people make use of correlated information for sustaining cooperation given that everyone reports truthfully. In other words, our design does not directly answer the research question of Awaya & Krishna (2016; 2019), that is, what effect the exchange of private information has on cooperation when signals are correlated. Instead, we experimentally test the underpinnings of their theory by addressing the question to what extent subjects can exploit the information about others' actions contained in the correlation of signals.[2]

For this purpose, we implement an indefinitely repeated Prisoners' Dilemma of imperfect public monitoring, where payoffs depend on the own action and the signal regarding the action of the other player. Signals are publicly observable and noisy, that is they falsely represent the real action of the opponent with an exogenously given and fixed probability, which is known to subjects. We distinguish between two settings. In one experimental setting we implement a correlation structure which can be exploited to support full cooperation in a public perfect equilibrium. Signals are systematically and perfectly correlated, that is the two public signals are the same, if both players choose the same action. If their actions differ, signals are independently drawn. We compare subjects' decisions in this setting (*Cor*) to a control setting with independent signals using the same stage game parameters and continuation probability (*No Cor*). The control setting without correlation has a similar design as Fudenberg et al. (2012), Aoyagi et al. (2019) and Dvorak and Fehrler (2024).[3]

[1]Private monitoring, as in Sekiguchi (1997); Compte & Postlewaite (2015), means that noisy signals are only privately observable. The secret price cutting in oligopolies described by Stigler (1964) is a classic example.

[2]Findings from other studies of different contexts suggest that subjects may struggle to make correct inferences from correlated information (e.g., Enke & Zimmermann, 2019; Fehrler et al., 2026)

[3]Other recent experimental literature on repeated games includes Dal Bó & Fréchette (2019), Romero & Rosokha (2023), Aoyagi et al. (2024), Backhaus & Breitmoser (2026), and Bland (2025). For a survey of earlier contributions see Dal Bó & Fréchette (2018).

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Intuitively, if publicly observable signals are perfectly correlated, that is identical when actions are the same and independently drawn otherwise, players can look for differences in signals to maintain cooperation. In such cases, simple cooperative strategies can be used that condition on the correlation of the signals. For instance, cooperation can be enforced by a grim-trigger strategy which triggers if signals diverge. If both players start with cooperation, both will receive the same signal, and will continue cooperating as long as the two signals match. Once the signals become different, they infer from the dissimilarity of the signals that different actions have been chosen in this period. The cooperating player therefore detects the defection of the other player with certainty and defects in all future periods. As this strategy is similar to the grim-trigger strategy under perfect monitoring except that it takes the correlation of information into account, we refer to this strategy as the correlated (signals) grim-trigger strategy (CGRIM). We also consider other correlation-based strategies such as correlated tit-for-tat (CTFT), correlated win-stay-lose-shift and a correlated trigger strategy with two periods of punishment (CT2). In the experiment, we use a combination of stage-game parameters and continuation probability for which the perfect correlation structure described above allows full cooperation based on CGRIM, while at the same time no cooperative public perfect equilibria exist in the treatment without signal correlation.

Our design includes treatments both with and without pre-play communication. When pre-play communication is allowed, it takes the form of free chat before round one of each indefinitely repeated interaction. Communication facilitates cooperation (e.g., Cooper et al. 1992; Rabin 1994; Ellingsen & Östling 2010). Pre-play communication, in particular, helps players to coordinate on cooperative equilibria, and raises cooperation rates by reducing strategic uncertainty (Kartal & Müller, 2026). In a previous study of a noisy, indefinitely repeated Prisoners' Dilemmas with uncorrelated signals, Dvorak and Fehrler (2024) observe high cooperation rates in the first round with pre-play communication followed by a steady and substantial decline in the subsequent rounds. We run treatments with pre-play communication to have a high chance of observing cooperation in the beginning of an interaction, and to be able to compare the decline in cooperation rates over rounds between treatments. Furthermore, we categorize communication content and analyze its impact on behavior.

Our main finding is that more than half of all subjects recognize the value of the correlation structure for sustaining cooperation and adopt correlation-based strategies when signals are perfectly correlated and pre-play communication is allowed. In the absence of pre-play communication, 17% of strategies are still correlation-based when signals are perfectly correlated. By contrast, no correlation-based strategies are observed when signals are not correlated, regardless of whether pre-play communication is permitted. With pre-play communication, the majority of these subjects use the CGRIM strategy, while a few also play CTFT. Without pre-play communication, only CGRIM is used. However, under both communication conditions we find no evidence that the correlation structure of signals promotes cooperation. Correlation plays two opposing roles: it makes defection detectable with certainty, which in principle enables efficient cooperation, but that same certainty makes responses to bad signals unforgiving, so the non-lenient CGRIM triggers frequent punishment. With pre-play communication, this loss of leniency in *Cor* is offset by the high leniency subjects display in *NoCor*, where defection cannot be detected with certainty; without communication, cooperation is substantially lower in both conditions.

The rest of the paper is organized as follows. In the next section, we introduce the stage-game, and derive theoretical predictions for both treatments. In Section 3, we describe the experimental design and explain the reasons for our choice of stage-game parameters. Section 4 summarizes our research questions and methods to address them. We present the empirical results in Section 5. The key findings are summarized and discussed in Section 6.

2. Repeated Prisoners' Dilemma with imperfect public monitoring

Consider an indefinitely repeated Prisoners' Dilemma with two players who repeatedly play a 2×2 stage game. The game terminates with a constant exogenously given probability of $0 < 1 - \delta < 1$

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Table 1. The prisoner's dilemma stage game

	C	D
C	$1, 1$	$-l,\ 1+g$
D	$1+g,\ -l$	$0, 0$

after every round. In each round, the two players $i = \{1, 2\}$ choose from an action set $A_i = \{C, D\}$ and their action is denoted by a_i.

There is ex-post uncertainty regarding the actual choice of an opponent. Under public monitoring (Green & Porter, 1984), instead of observing the actual choice, each player's action is translated into a noisy signal and both signals are publicly announced to players. More technically, let $\Omega_i \in \{c, d\}$ be a set of signals of player i's action, where signal c represents action C, and d represents D. We denote the realized signal from this set by ω_i. For each action profile $a = (a_i, a_{-i})$, a conditional probability distribution $\pi(\omega|a)$ is assigned over signals. In our experiment, the two treatments differ with respect to π, meaning the quality of the signals differs as well.

Imperfect public monitoring with uncorrelated signals (*NoCor*) For the treatment without correlation, signals are conditionally independent of $a = (a_1, a_2)$. That is, signals are drawn independently for each of the chosen actions, and signal ω_i depends only on a_i but not on a_{-i}. Signals are noisy and indicate the opposite of the chosen action with probability $\epsilon = 0.2$, that is: when the player chooses C (D), the signal indicates C (D) with probability $1 - \epsilon = 0.8$, and indicates D (C) with probability 0.2. The conditional probability distribution of ω_i is thus $\pi(\omega_i = c|a_{-i} = C) = \pi(\omega_i = d|a_{-i} = D) = 1 - \epsilon$ and $\pi(\omega_i = c|a_{-i} = D) = \pi(\omega_i = d|a_{-i} = C) = \epsilon$.

Imperfect public monitoring with correlated signals (*Cor*) In this setup, both players receive identical signals if they choose the same action. Yet the signals still indicate the opposite of the chosen action with probability 0.2. The conditional probability distribution of $\omega_i (= \omega_{-i})$ is given by $\pi(\omega_i = c|a_i = a_{-i} = C) = \pi(\omega_i = d|a_i = a_{-i} = D) = 1 - \epsilon$ and $\pi(\omega_i = c|a_i = a_{-i} = D) = \pi(\omega_i = d|a_i = a_{-i} = C) = \epsilon$ when both players choose the same action. If their actions differ, signals are drawn independently, as in *NoCor*. Therefore, if the two signals differ, it is certain that different actions were chosen. If the signals are the same, players do not know for sure the opponent's action.

Player i's stage game payoff y_i is defined by i's action a_i and the noisy signal ω_{-i} reflecting the other player's action, that is $y_i : A_i \times \Omega_{-i} \to \mathbb{R}$. Hence player i cannot infer the action of the other player from her payoff. The expected stage game payoff of player i is given by $u_i : A \to \mathbb{R}$, $u_i(a) = \sum_{\omega_{-i} \in \Omega_{-i}} y_i(a_i, \omega_{-i})\pi(\omega_{-i}|a)$. The expected stage-game payoff profile (u_i, u_{-i}) which we consider has the form of a Prisoners' Dilemma. Expected payoffs in Table 1 are normalized. Parameters g and l are both positive and $g < 1 + l$.

Denote by $\Omega^t = \{\Omega_i, \Omega_{-i}\}^t$ the set of public histories up to round t. A public strategy for player i is a mapping $\sigma_i : \bigcup_{t \geq 0} \Omega^t \to \Delta A_i$. A strategy profile $\sigma = (\sigma_i, \sigma_{-i})$ is a perfect public equilibrium (PPE) if $\forall i : \sigma_i$ is a public strategy and for any history up to round t, σ is a Nash equilibrium in all rounds following t.

The equilibrium solution for *NoCor* Mutual cooperation can be sustained by non-lenient and non-forgiving grim-trigger strategies. Players start with C but deviate to D for all subsequent rounds if $(\omega_i, \omega_{-i})^t \neq (c, c)^t$. We refer to mutual cooperation as the reward state, and to mutual defection as the punishment state. This strategy is a PPE if the long-run incentive of cooperation $\frac{1}{1-\delta(1-\epsilon)^2}$ is as least as big as the gain from defection $\frac{1+g}{1-\delta\epsilon(1-\epsilon)}$, that is δ should be sufficiently large:

$$\delta \geq \delta_{NoCor}^{PPE} = \frac{g}{(1+g)(1-\epsilon)^2 + \epsilon^2 - \epsilon} \tag{1}$$

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

The equilibrium solution for *Cor* Mutual cooperation can still be enforced by grim-trigger strategies. The class of cooperative PPE strategies defined in the above paragraph still exists. In addition, correlated grim-trigger strategies exist which can be used to achieve mutual cooperation in all rounds. Specifically, we define the correlated grim-trigger strategy (CGRIM) as follows: the player starts with C and will continue choosing C as long as the two public signals match, that is $\omega_i = \omega_{-i}$. Otherwise the player chooses D in all subsequent rounds. This strategy can be used to construct a symmetric public perfect equilibrium if the long-term gains of cooperation are at least as big as the short-term incentive to defect. The threshold δ^{PPE} is given by

$$\delta \geq \delta_{Cor}^{PPE} = \frac{g}{1 - 2\epsilon + 2\epsilon^2 + g} \tag{2}$$

See Section A of the Supplementary Material for a proof.

Some remarks are in order. It is easier to sustain cooperation under *Cor* than *NoCor* in terms of grim-trigger strategies. For $\epsilon \in (0, 1)$ and g fixed, whenever (1) holds, condition (2) holds as well. The correlation structure of signals changes the equilibrium condition in two aspects.

First, it increases the value in the continuation path of the reward state. Even when both players cooperate, bad signals occur with positive probability in *NoCor* which triggers the punishment state. Under *NoCor*, the continuation value on the cooperative path is the expected utility resulting from being in the reward or punishment state, both of which will be reached with positive probability. However, with correlation, signals always match if both cooperate. This means that the punishment state will never be reached and the continuation value will be larger.

Second, note that unilateral defection is more attractive in *NoCor* than in *Cor*. If the signals differ in *Cor*, a cooperating player can be certain that their opponent has defected. However, it is impossible to make such an inference when signals match. Even when players choose different actions, signals still match with a probability of $2\epsilon(1-\epsilon)$. This is the probability of remaining in the reward state after unilateral defection, which corresponds to the probability of the signal profile being either $(\omega_i, \omega_{-i}) = (c, c)$ or (d, d). The *NoCor* treatment renders this probability to be $\epsilon(1 - \epsilon)$, that is, cooperation in the next round is only possible when (c, c) is observed. *NoCor* triggers punishment with a higher probability when one player defects unilaterally. However, it does not compensate for the reduced expected utility in the reward state due to punishments on the equilibrium path.[4]

The equilibrium selection criterion based on the discount factor threshold discussed above may be too coarse: it asks only whether CGRIM can be sustained as an equilibrium, but says nothing about how attractive CGRIM is relative to always defecting (ALLD). To capture this, in Section B of the Supplementary Material we analyze a 2×2 game in *Cor* in which players are restricted to either CGRIM or ALLD. CGRIM is risk dominant when it yields a higher payoff than ALLD against an opponent who plays CGRIM and ALLD with equal probability.

3. Experimental design

We implement a noisy Prisoners' Dilemma game with public monitoring in a laboratory experiment. The experiment has four between-subject treatments. The four treatments vary in the conditional distribution of signals and in whether pre-play communication is possible.[5]

NoCor-Chat Signals are public and independent. Pre-play communication is possible.

[4]The difference in the probability to stay in reward state when choosing defection instead of cooperation is larger in *Cor* ($1 - 0.32 = 0.68$) than in *NoCor* ($0.64 - 0.16 = 0.48$). Therefore, unilateral defection is more risky in *Cor* compared to *NoCor*. Since the reward state is also more valuable in *Cor*, unilateral defection is less attractive in *Cor*.

[5]The two treatments with pre-play communication have been pre-registered on the AEA RCT Registry (Bao et al., 2020, AEARCTR-0005369). We also present our pre-analysis plan in Section I of the Supplementary Material. The two treatments without pre-play communication were requested during the reviewing process and therefore do not appear in the pre-registration.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Cor-Chat Signals are public and perfectly correlated if both actions are the same, and independent otherwise. Pre-play communication is possible.
NoCor-NoChat Signals are public and independent. Pre-play communication is not possible.
Cor-NoChat Signals are public and perfectly correlated if both actions are the same, and independent otherwise. Pre-play communication is not possible.

In every round, two players choose their actions $a_i \in \{C, D\}$ simultaneously. Payoffs depend on the player's own action a_i and the received signal about the other player's action $\omega_{-i} \in \{c, d\}$. Signals are noisy and indicate the wrong action with probability $\epsilon = 0.2$, which does not vary between treatments. This error probability is known to subjects.

The continuation probability δ of the repeated game is 0.8. The stage-game payoffs for each player are given in the upper panel of Figure 1. The payoffs are in experimental currency units. The lower panel of Figure 1 shows the expected stage-game payoffs for action profiles of both players. The normalized expected stage-game parameters are $g = l = 0.8$.

Under this parameterization, the existence thresholds for cooperative PPEs are $\delta^{PPE}_{NoCor} = 0.81$ and $\delta^{PPE}_{Cor} = 0.54$. These parameters make sure that the condition outlined in Equation (2) holds but the one defined in Equation (1) does not; that is, there exists an equilibrium in which both players play correlated grim-trigger strategy when signals are correlated, and no cooperative PPE exists if there is no correlation.[6]

Under this parameterization, our discount factor also exceeds the risk-dominance threshold derived in Section B of the Supplementary Material ($\delta = 0.8 > 0.702$), so CGRIM is not only sustainable as an equilibrium but also risk dominant. This further strengthens the theoretical prediction of cooperation in *Cor*.

In the treatments with pre-play communication, subjects engage in a communication-stage before the first round of every supergame. A supergame is one indefinitely repeated interaction. Communication is possible via a chat-box interface for 120 seconds.

We include pre-play communication because our central question is whether cooperation, once established, can be *sustained* by exploiting correlated signals. Testing this requires cooperation rates that start high enough for differences in sustainability to be detectable; without a coordination device—especially in *NoCor*, where no cooperative equilibrium exists—initial cooperation may be too low for such a test. Free-form pre-play communication reliably raises initial cooperation (Dvorak & Fehrler, 2024), providing a common, high starting point from which we can compare the stability of cooperation across treatments. The *NoChat* treatments, added during the review process, complement this design by providing another test of the correlation effect without the potential confound of communication.

In every session of all treatments, subjects are randomly assigned to 3 matching groups with at least 8 participants in each group. Subjects only interact with the members in the same matching group throughout the entire session. Therefore, the matching groups can be treated as independent observations.[7] Subjects play 7 supergames of pre-determined lengths. At the beginning of every supergame, subjects are matched with a new partner from their matching group using perfect stranger matching

[6]While more sophisticated strategies with finite punishment phases might achieve higher payoffs than grim trigger when cooperation is sustainable (e.g., Abreu et al., 1990), it is nevertheless true that if cooperation cannot be sustained by grim trigger, it cannot be sustained in any PPE. The spread between reward and punishment state values, $V_C - V_P$, which determines incentive power, is strictly decreasing in V_P: any forgiveness raises the punishment value directly but raises the cooperative value only indirectly through possible transitions to punishment. Since the indirect effect is strictly smaller, the spread shrinks. Therefore, if the incentive constraint fails under grim trigger ($V_P = 0$, with normalized stage game pay-offs), it must fail for any $V_P > 0$.

[7]Our ex-ante power analysis, documented in the pre-analysis plan (see Section I of the Supplementary Material), indicated that 9 matching groups per treatment provide sufficient statistical power ($> 80\%$) to detect effect sizes of 10 percentage points based on 10,000 simulation iterations.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

	c	d
C	32	2
D	40	10

	C	D
C	$26, 26$	$8, 34$
D	$34, 8$	$16, 16$

Figure 1. Stage-game pay-offs

Notes: Payoffs are in experimental currency units. The exchange rate is 50 ECU = 1 EUR for the two treatments with pre-play communication, and 20 ECU = 1 EUR for the two treatments without pre-play communication. Both matrices are shown to subjects during their decision-making.

such that they do not play with the same partner from the matching group for a second time. To keep the length of supergames constant across treatments, we generate 3 sequences of random numbers beforehand, and use them to determine the length L_j of each supergame j.[8] To increase the number of observations per supergame, we adapt the block-random-termination method (Fréchette & Yuksel, 2017). Subjects play a block of five rounds at the beginning of every supergame. If the true length L_j is smaller or equal to 5, the supergame ends at the end of round 5 and only the first L_j rounds are payoff relevant. If L_j is larger than 5, the supergame continues until round L_j has been reached and all rounds are payoff relevant. Before the end of round 5, subjects are not informed about whether the supergame ends or not. The modified block-random-termination method allows us to collect data of at least five rounds in every supergame.

At the end of each of the first four rounds within every supergame, subjects receive feedback on $\{a_i, \omega_i, \omega_{-i}\}$ plus the stage profit y_i. Starting from round 5, in addition to $\{a_i, \omega_i, \omega_{-i}\}$ and y_i, they are informed of the pre-generated random number of that round. The supergame does not end until that number is smaller or equal than $1 - \delta$, which is 0.2 under our parameterization.

Subjects are provided with a detailed explanation of the experimental procedure, and after reading these instructions they must answer a set of comprehension (control) questions to verify their understanding before the game begins (see Section H of the Supplementary Material). At the end of the experiment, the subjects answer a short survey. The survey asks about basic socioeconomic characteristics, such as age and gender, as well as how well the subjects understood the instructions.

All sessions of the experiment were run at Lakelab, University of Konstanz. Six sessions of the *Chat* treatments were conducted in February 2020, with three sessions each of *Cor-Chat* and *NoCor-Chat*. Six sessions of the *NoChat* treatments followed in November 2025, again with three sessions each of *Cor-NoChat* and *NoCor-NoChat*. A total of 144 subjects participated in our experiment in 2020, and 172 participated in 2025.[9] Subjects were students of the University of Konstanz.[10] Each experimental

[8] We use Stata to generate 3 sequences of uniformly distributed random numbers between 0 and 1 with seeds 3, 4, and 5 (Seeds 1 and 2 have been used in: Dvorak & Fehrler, 2024). Denote the 3 sequences as $\{r_n\}_i = \{r_1, r_2, \ldots, r_x\}_i$, where $i = 3, 4, 5$ indicates the seed underlying the sequence and $n \in \mathbb{N}$. The first supergame has x_1 rounds if $r_{x_1} \leq 0.2$ and for all $n < x_1, r_n > 0.2$. The second supergame has $x_2 - x_1$ rounds if $r_{x_2} \leq 0.2$ and for all $x_1 < n < x_2, r_n > 0.2$. And so forth. The resulting (lengths of the) sequences are SQ1 $(2, 8, 1, 5, 7, 1, 7)$, SQ2 $(4, 2, 2, 21, 4, 3, 5)$ and SQ3 $(2, 3, 1, 1, 4, 6, 6)$.

[9] After data collection, we found out that one subject participated twice in two *NoCor-Chat* sessions. For the *Chat* treatments, we exclude the matching group of this subject's second participation. The analysis is based on the remaining 136 subjects, with a mean age of 22 years (54.41% female). Gender is balanced between treatments ($p = 0.86$). Results on strategies and cooperation rates including all matching groups are in Section D of the Supplementary Material. Including all matching groups yields results very similar to those reported in the main manuscript. For the *NoChat* treatments, subjects had a mean age of 21 years, and 66.86% were female. Gender distribution does not differ significantly between treatments ($p = 0.94$).

[10] The experiment was programmed in z-Tree (Fischbacher, 2007) and subjects were recruited via hroot (Bock et al., 2014).

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

session lasted approximately two hours. Average earnings were 18.67 EUR in 2020 and 30.73 EUR in 2025.

4. Research questions & methods

We address three research questions.

Question 1. *Does correlation in signals increase cooperation?*

According to the theoretical prediction outlined in Section 2, there should be more cooperation in the *Cor* treatments. In line with the theory, our pre-registered main hypothesis is the following:
H1: *The average cooperation rate will be higher in Cor than in NoCor.*

H1 follows directly from the critical continuation probabilities defined in Equations (1) and (2), and our stage-game parameters ($g = l = \delta = 0.8$). We test this hypothesis by comparing the average cooperation rates in the first five rounds of the last three supergames between treatments. We further analyze the decline in cooperation rates to compare the stability of cooperation.

Question 2. *Do subjects choose different strategies when cooperation can be sustained via correlated signals?*

We start with an analysis of the cooperation rates after memory-one histories to shed light on treatment differences in leniency and forgiveness. In our setup, a memory-one history consists of a player's own action and the two public signals $\{a_i,\omega_{-i},\omega_i\}$. We have a vector of nine possible memory-one states ($\emptyset$, *ccc*, *ccd*, *cdc*, *cdd*, *dcc*, *dcd*, *ddc*, *ddd*). The first element $\emptyset$ is the initial round with no history of play. The rest of the elements are nonempty histories representing $\{a_i,\omega_{-i},\omega_i\}$. For instance, *cdc* describes a state of a player who chose cooperation, receives signal *d* and sends out signal *c* to the other player. We look at the cooperation probabilities after each of the possible histories and we represent the cooperation probabilities by a vector ($\sigma_\emptyset$, σ_{ccc}, σ_{ccd}, σ_{cdc}, σ_{cdd}, σ_{dcc}, σ_{dcd}, σ_{ddc}, σ_{ddd}). We focus on nine states to allow behavior to be conditioned on the combination of action and public signals. This allows us to investigate whether the cooperation rates are in line with correlation-based strategies. Unjustified defection is indicated by $1 - \sigma_{ccc}$ without signal correlation, and by $1 - \sigma_{ccc}$ and $1 - \sigma_{cdd}$ with signal correlation. σ_{cdc}, σ_{cdd} can be interpreted as measures for leniency when signals are uncorrelated. When signals are correlated, measures of leniency are σ_{ccd} and σ_{cdc}. For both uncorrelated and correlated signals, σ_{dcc}, σ_{dcd}, σ_{ddc} and σ_{ddd} are measures of forgiveness.

To analyze strategy choices, we use the R package stratEst (Dvorak, 2022), which can be used to implement the SFEM of Dal Bó and Fréchette (2011) based on the EM algorithm (Dempster et al., 1977). We restrict our attention to the first five rounds of the last 3 supergames when subjects have gained experience of play. In the main text, we report models which include the 6 pure strategies studied by Dal Bó and Fréchette (2011) and their corresponding correlation-based variants as our candidate strategies. As this set of strategies was originally developed for the case of perfect monitoring, we adapt these pure strategies so that they depend on own actions and public signals. A summary of the candidate strategies is provided in Table 2. As a robustness check, we extend the candidate set to include the 20 strategies analyzed in Fudenberg et al. (2012) for the imperfect public monitoring case. Descriptions of each strategy can be found in Section E of the Supplementary Material.

Question 3. *How does communication affect cooperation?*

To approach this question, we had two research assistants classify the subjects' communication content into 39 sub-categories, which we then grouped into four main categories. Table G1 in the

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Table 2. Overview of the candidate strategies

Strategy	Acronym	Description
Always Defect	ALLD	Always play *D*.
Always Cooperate	ALLC	Always play *C*.
Grim	GRIM	Play *C* until either signal is *d*, then play *D* forever.
Tit-for-Tat	TFT	Play *C* unless the signal representing the partner's action is *d* in the last round.
Win-Stay-Lose-Shift	WSLS	Play *C* if own choice is the same as the signal representing the patner's choice in the last round, otherwise play *D*.
T2	T2	Play *C* until either signal is *d*, then play *D* twice and return to *C* (regardless of all actions and signals during the punishment rounds).
Correlated Grim	CGRIM	Play *C* until the public signals do not match, then play *D* forever.
Correlated Tit-for-Tat	CTFT	Play *C* unless the public signals do not match in the last round.
Correlated Win-Stay-Lose-Shift	CWSLS	Play *C* if the public signals match, otherwise play *D*.
Correlated T2	CT2	Play *C* until the signals do not match, then play *D* twice and return to *C* (regardless of all actions and signals during the punishment rounds).

Supplementary Material presents the categories. We expect that subjects who coordinate in the *Cor-Chat* treatment, explicitly or implicitly talk about punishments for the case when the two public signals differ. Our primary focus is on the main categorical level. We treat these as explanatory variables and use logistic regression to analyze whether subjects who discuss certain topics are more likely to play cooperatively.

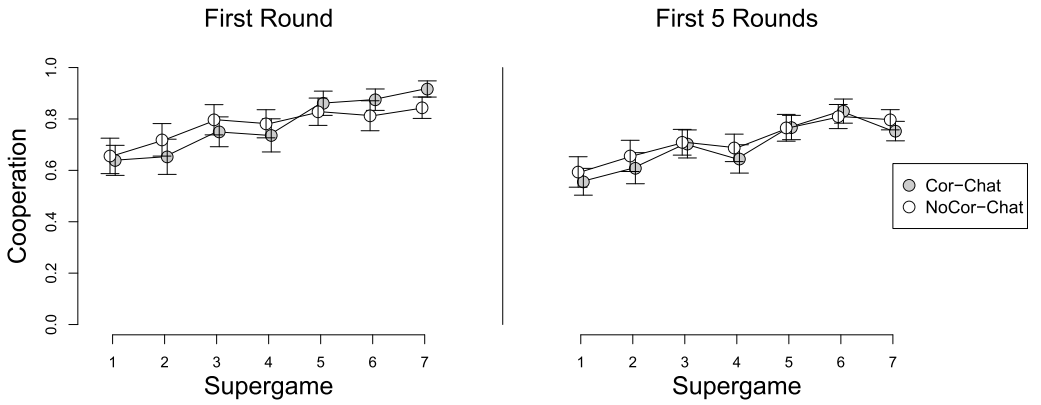
5. Experimental results

In line with our pre-analysis plan, we focus on the last three supergames for cooperation rates and strategy frequency analysis, when subjects have gained experience with the game. Figure 2 depicts the average cooperation frequency over all seven supergames. The cooperation rates are stable over the supergames in treatments without communication. In the two treatments with communication, they increase as subjects gain experience. In addition, since each supergame consists of at least five rounds, our main focus is on the first five rounds in the supergames such that the transition from the block to random termination does not confound the results.

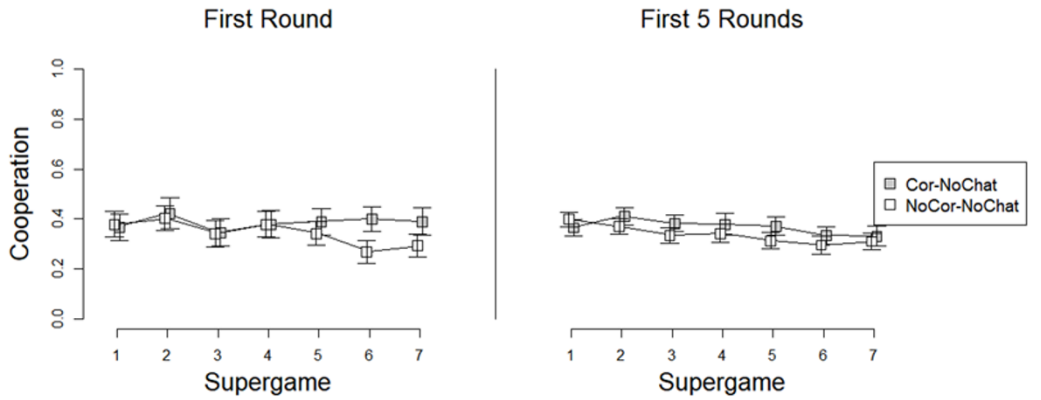
5.1. Cooperation

Figure 3 shows the average cooperation rate with and without correlation under the two communication conditions. Bars indicate the means, and error bars indicate two-way clustered standard errors of the mean (Cameron et al., 2011). The two cluster variables we use for the standard errors are participants and matches. Figure 3 indicates that the average cooperation rate does not differ between the *Cor* and *NoCor* conditions, irrespective of whether communication is allowed. With pre-play communication, the average cooperation rate is 0.78 in *Cor-Chat*, 0.79 in *NoCor-Chat*. Treating each matching group as an independent observation, the difference in cooperation frequency is not statistically significant (one-sided t-test, $p = 0.55$). Without pre-play communication, the average cooperation rates are 0.35 and 0.31 for *Cor-NoChat* and *NoCor-NoChat* respectively. Treating each matching group as an independent observation, the one-sided t-test shows that the cooperation rates are not different between treatments ($p = 0.25$). A subject-level regression analysis with two-way clustered standard errors, reported in Table C1 in the Supplementary Material, confirms

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press



(a) With Pre-play Communication



(b) Without Pre-play Communication

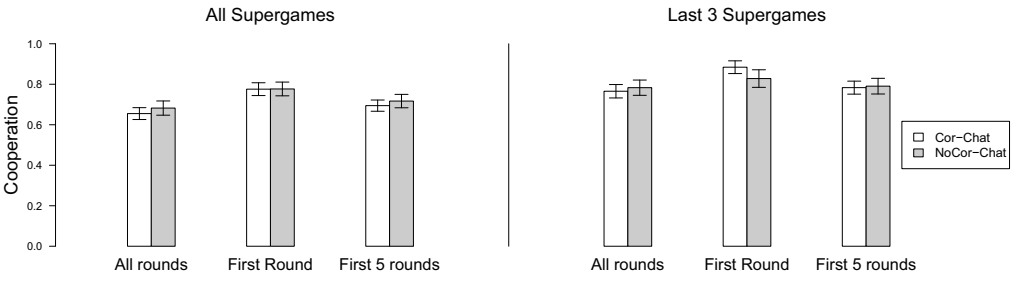
Figure 2. Evolution of cooperation over supergames
Notes: The lines depict the first round and average frequency of cooperation over seven supergames for the four treatments. Error bars represent two-way clustered standard errors for the mean cooperation rate on subject and match level.

these findings; in particular, the effect of correlation remains statistically insignificant even when the analysis is restricted to first-round choices in the last three supergames.[11]

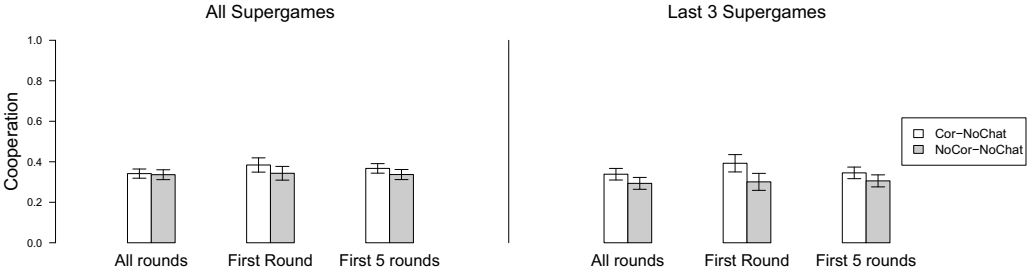
Figure 4 shows the average frequency of cooperation over the first five rounds. With communication, the cooperation rates start high (above 78 percent in all supergames, and above 82 percent in the last three supergames) and decline over rounds. In the last three supergames, the average cooperation rate has reduced by 19 percentage points in *Cor-Chat* after five rounds and by 9 percentage points in *NoCor-Chat*, which indicates a higher stability in *NoCor-Chat*. We regress subject-level cooperation choices on the round number, supergame number and treatment dummies to test whether the difference in stability is statistically significant using two-way clustered standard errors on subject and match. The result shows that the decline in cooperation is steeper in *Cor-Chat* than in *NoCor-Chat* ($p = 0.04$). The greater stability observed in *NoCor-Chat* is likely driven by higher leniency. As shown in Tables 3 and 4, with correlation, subjects respond less leniently following the state σ_{cdc},

[11] For all four treatments, we asked subjects in a post-experimental questionnaire whether they understood the instructions. Approximately 83% of subjects reported that they did. Regression results indicate that self-reported understanding is not associated with higher cooperation rates ($p = 0.83$). Restricting the sample to the two treatments with correlation yields the same conclusion ($p = 0.66$).

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press



(a) With Pre-play Communication



(b) Without Pre-play Communication

Figure 3. Average cooperation rate between treatments
Notes: Bars show the average frequency of cooperation. Error bars represent two-way clustered standard errors for the mean cooperation rate on subject and match level.

and they adopt the unforgiving punishment strategy CGRIM. When using the aggregated matching-group averages, the decline in cooperation in *Cor-Chat* is not statistically significantly steeper than in *NoCor-Chat* ($p = 0.26$).

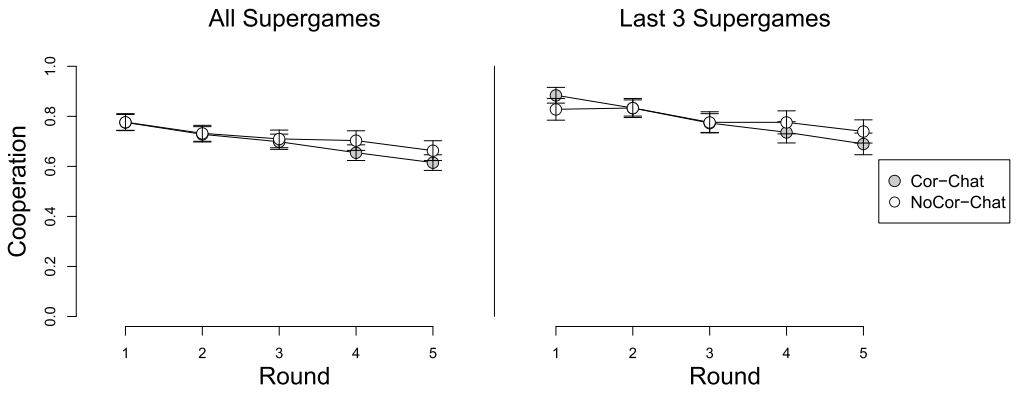
When pre-play communication is not involved, the average frequency of cooperation is stable over the first five rounds. When focusing on the last three supergames, the cooperation rates decline by approximately 10 percentage points in *Cor-NoChat*, and less than 1 percentage point in *NoCor-NoChat*. Regression results on the subject level show that the decline in cooperation is not statistically significantly steeper in *Cor-NoChat* than in *NoCor-NoChat* ($p = 0.07$). One possible explanation is that the initial cooperation is already low, leaving little scope for the defection-detection effect of the correlation device to trigger punishment. When treating each matching group as an independent observation, the decline in cooperation is also not statistically significantly different between the two treatments ($p = 0.15$).

Result 1: *Signal correlation does not increase the cooperation rate, regardless of whether pre-play communication is available. With pre-play communication, the decline in cooperation is slightly steeper in the correlation treatment than in the no-correlation treatment.*

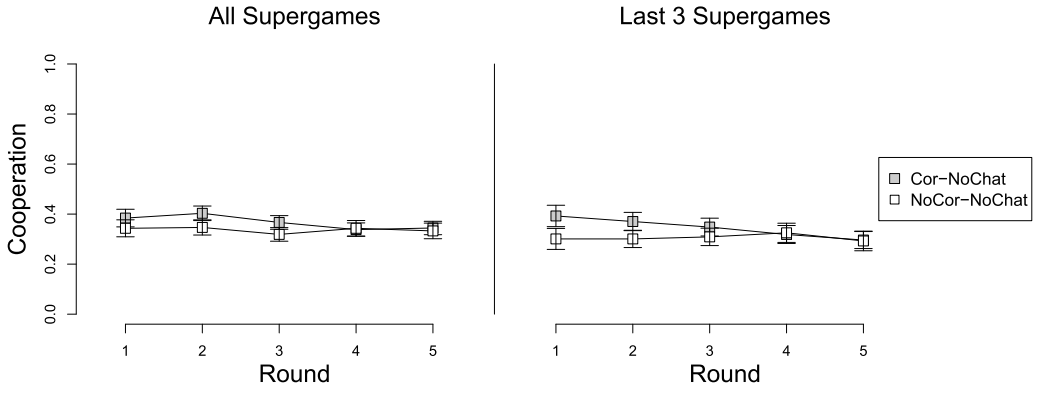
5.2. Strategy choices

We estimate the probability of cooperation following different memory-one histories. For both treatments, the cooperation rate may depend on the set of nine possible memory-one histories, denoted by $(a_i, \omega_{-i}, \omega_i)$. We first discuss the results in treatments with pre-play communication. $\sigma_\emptyset$ indicates the probability of cooperating in the first round. It is high in treatments *Cor-Chat* and *NoCor-Chat*,

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press



(a) With Pre-play Communication



(b) Without Pre-play Communication

Figure 4. Stability of cooperation over rounds

Notes: The lines show the average frequency of cooperation in the first 5 rounds of all and the last 3 supergames. Error bars represent two-way clustered standard errors for the mean cooperation rate on subject and match level.

Table 3. Cooperation rates after memory-one histories

	$\sigma_{\emptyset}$	σ_{ccc}	σ_{ccd}	σ_{cdc}	σ_{cdd}	σ_{dcc}	σ_{dcd}	σ_{ddc}	σ_{ddd}	lnL
With Pre-play Communication										
Cor	0.88	0.95	0.83	0.33	0.71	0.41	0.39	0.17	0.35	-409.28
	(0.03)	(0.01)	(0.22)	(0.06)	(0.04)	(0.08)	(0.09)	(0.21)	(0.07)	
NoCor	0.83	0.91	0.95	0.79	0.86	0.22	0.46	0.11	0.32	-386.73
	(0.04)	(0.02)	(0.03)	(0.04)	(0.07)	(0.13)	(0.10)	(0.09)	(0.09)	
Without Pre-play Communication										
Cor	0.39	0.37	0.20	0.33	0.42	0.29	0.37	0.40	0.31	-864.68
	(0.04)	(0.06)	(0.13)	(0.04)	(0.07)	(0.05)	(0.05)	(0.16)	(0.03)	
NoCor	0.33	0.27	0.35	0.26	0.28	0.20	0.35	0.25	0.34	-752.08
	(0.04)	(0.05)	(0.17)	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)	(0.04)	

Notes: This table summarizes the response probabilities of cooperation following the 9 possible histories: $\emptyset$, ccc, ccd, cdc, cdd, dcc, dcd, ddc, ddd. Bootstrapped standard errors are from 10000 iterations and are presented in parentheses. The log likelihood of the model is summarized in the last column.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Table 4. SFEM with strategy selection

	ALLD	ALLC	GRIM	TFT	WSLS	CGRIM	CTFT	γ	BIC	ln*L*
With Pre-play Communication										
Cor	0.06 (0.03)	0.26 (0.08)	-	0.06 (0.04)	0.07 (0.05)	0.40 (0.10)	0.16 (0.08)	0.10	856.59	-415.47
NoCor	0.12 (0.04)	0.76 (0.06)	-	0.07 (0.04)	-	0.05 (0.03)	-	0.10	737.02	-360.19
Without Pre-play Communication										
Cor	0.64 (0.06)	0.19 (0.05)	-	-	-	0.17 (0.05)	-	0.24	1604.52	-795.51
NoCor	0.74 (0.05)	0.19 (0.05)	0.07 (0.04)	-	-	-	-	0.20	1338.21	-662.49

Notes: The table reports the maximum-likelihood shares of the strategies of Dal Bó and Fréchette (2011) and their correlation-variants with data from the first 5 rounds of the last 3 supergames. All strategies condition on the history of public signals. The estimation procedure assumes constant strategy use. γ is the estimated tremble probability, which avoids likelihood shares of zero when subjects deviate from a choice pattern. Strategies are selected based on Bayesian Information Criterion. Strategies attracting zero shares are omitted (-). Standard errors are reported in parentheses. Values may not add up to one because of rounding.

meaning that subjects are generally cooperative in the initial round. The unjustified defection ($1-\sigma_{ccc}$, $1-\sigma_{cdd}$ in *Cor-Chat* and $1-\sigma_{ccc}$ in *NoCor-Chat*) rates are low in both treatments. σ_{cdc} and σ_{ccd} can be interpreted as measures of leniency in both treatments. A higher level indicates that subjects are less likely to defect when a bad signal occurs. The largest treatment difference is observed after the *cdc* history. Subjects in *NoCor-Chat* are substantially more cooperative than in *Cor-Chat*. Treatment differences in σ_{ccd} are small. We do not observe much difference in σ_{dcc} and σ_{dcd} either. These cooperation rates are relatively low in both treatments. In terms of forgiveness (σ_{dcc}, σ_{dcd}. σ_{ddc} and σ_{ddd}), we see that the willingness to return to cooperation after defection is low in both *Cor-Chat* and *NoCor-Chat*.

Without pre-play communication, subjects are much less cooperative in the initial round and overall. The unjustified defection rates are lower than the justified defection rates and subjects in *Cor-NoChat* are more lenient toward a defective signal regarding their partner's action as long as both signals match. Some of them appear to make use of the correlation structure and adjust their behavior accordingly.

Behavior is more lenient in the treatments with pre-play communication. After the history *cdc*, the high cooperation rate of 0.79 in the no-correlation condition decreases to 0.26 when communication is absent. Similarly, after history *ccd*, the cooperation rate in the correlation condition decreases from 0.83 to 0.20.

The probabilities of reaching these histories differ for each strategy. Therefore, we cannot interpret the cooperation probabilities after memory-one histories as strategy choices. For instance, σ_{ccd} and σ_{cdc} are very different in the correlation treatment with pre-play communication. Subjects are willing to cooperate when they receive a cooperative signal from their partners but tend to defect when the signal regarding the partner's choice is defective. If subjects play CGRIM, σ_{ccd} and σ_{cdc} should be equal. But we still cannot conclude that subjects in *Cor* do not play CGRIM due to the above reason. To find out what strategies they play, our next step is to investigate the heterogeneity of strategy choices. We are interested in the specific strategies that subjects play, which help to explain our observations in Table 3 in terms of why subjects in *NoCor-Chat* are substantially more lenient, on average, following *cdc*, and substantially less lenient following *cdd* without communication. All candidate strategies condition on $\{a_i, \omega_{-i}, \omega_i\}$.

We fit a strategy estimation model on the candidate strategy set and select the subset of strategies that best explains subjects' choices using the Bayesian Information Criterion (BIC) for model

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

comparison, via stepwise backward elimination.[12] Table 4 shows the estimated maximum-likelihood shares of the selected strategies for both correlation treatments under different communication scenarios. To assess whether strategies differ in two treatments, we fit a model on the pooled data and block-bootstrap the likelihood-ratio test statistic by resampling individuals with replacement. For the block bootstrap, we use all the decisions from the same individual if that individual is sampled to account for dependency between decisions from the same individual. Estimating a restricted model with a smaller candidate set of strategies that excludes the correlation variants allows us to test whether the estimated strategy shares differ across candidate sets. The p-value of this test is < 0.001 with pre-play communication and 0.09 without pre-play communication. This indicates that excluding correlation-variants affects the estimated strategy shares only in the presence of pre-play communication.

Taking a closer look at the two treatments with pre-play communication, Table 4 shows that ALLC attracts a substantial share in *NoCor-Chat*, whereas the non-lenient and non-forgiving correlation-based variant of the grim-trigger strategy, CGRIM attracts the highest share in *Cor-Chat*. This estimation result explains and confirms the low leniency of subjects choices in *Cor-Chat* following state *cdc*. The behavior of a substantial share of subjects in *Cor-Chat* is consistent with playing CGRIM. They defect immediately when signals mismatch and never return to cooperation.

Without pre-play communication, ALLC, ALLD, and CGRIM attract non-zero shares in *Cor-NoChat*. Subjects play ALLC, ALLD, and GRIM in *NoCor-NoChat*. Among these strategies, always-defecting is the most popular one. Similar to the no-communication treatments, CGRIM is only played in the *Cor* condition. This explains why subjects are more lenient following *cdd* in *Cor-NoChat*, because for those who are playing CGRIM, defection is only triggered when the signals do not match.

The estimated ALLC share of 76% in *NoCor-Chat* is notably higher than what is typically found in the literature. Studies using the SFEM in repeated Prisoner's Dilemma games without communication consistently find that conditionally cooperative strategies such as Tit-for-Tat and Grim Trigger dominate, with ALLC attracting negligible shares (Dal Bó & Fréchette, 2011; Fudenberg et al., 2012; Breitmoser, 2015). The high ALLC share in our setting likely reflects the combination of free-form communication, which reduces strategic uncertainty, and uncorrelated signals, which limit the scope for conditional punishment. Accordingly, ALLC drops to 26% in *Cor-Chat*, where defection detection is feasible, and to 19% in *NoCor-NoChat*, where communication is absent.[13]

Result 2: *Strategy choices differ when signals are correlated. Subjects move from lenient strategies used in the absence of correlation to non-lenient and unforgiving correlation-based strategies with correlation.*

5.3. Communication

Next, we take a closer look at the pre-play communication content and explore whether and in what way the average cooperation rate and strategy choices are influenced by the opportunity to communicate. Two research assistants classified subject-round communication observations into 39 sub-categories, with 5.67 classifications on average for each subject-round observation and 2313

[12]The selection procedure works as follows. We first fit the model containing all K strategies. We then iteratively eliminate strategies: we calculate the BIC of all $K - 1$ nested models and select the best model in terms of BIC, repeating this process until none of the nested models improves BIC.

[13]To check the robustness of the SFEM results, we expanded the candidate strategy set to include another 14 pure strategies. Moreover, we include CWSLS and CT2 into the candidate set as alternative correlation-based strategies. Table F1 in the Supplementary Material presents the results. In general, ALLC remains the most popular strategy in *NoCor-Chat*, while in *Cor-Chat*, although a lenient variant of the grim-trigger strategy (GRIM2) is now the most prevalent strategy, CGRIM and CTFT attract significant shares. However, in treatments without pre-play-communication, the correlation-based strategies do not appear anymore. ALLD remains the most popular strategy. The lenient strategy GRIM2 is played in *Cor-NoChat*, whereas the non-lenient defective strategy DTFT is played in *NoCor-NoChat*.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Table 5. Frequency of valid codings per individual-round observation

	Cor-Chat	*NoCor-Chat*
Coordination	0.99	0.99
Deliberation	0.37	0.26
Relationship	0.15	0.03
Trivia	0.89	0.96

Notes: Frequency of valid codings on main categorical level of subject-round observations. Data is from the last three supergames. A classification is considered to be valid if the classifications from the two raters agree. It is possible that an observation belongs to more than one category, resulting in both column sums to be larger than 1.

classifications in total for the last three supergames. We merge the 39 sub-categories into four main categories: *Coordination*, *Deliberation*, *Relationship* and *Trivia*.[14] We determine Cohen's κ on the main categorical level.[15] The average κ is 0.35, indicating a fair level of agreement between the raters. *Coordination* is the effort to coordinate on future behavior. It includes explicit and implicit punishment threats when signals differ. According to Table 5, for both treatments, the category *Coordination* has been covered in almost all pre-play chats in the last three supergames. Observations concerning the discussion of actions and strategies are classified into the *Deliberation* category. The *Relationship* category contains all chats with promises, expressions of trust and distrust, and requests for trustworthy behavior. This category is the least frequent among the four. Another frequently observed category is *Trivia*. This category covers small talk, off-topic talks, and expressions of boredom.

A closer look at the sub-categorical levels (Table G1 in the Supplementary Material) shows that most subjects propose the mutual play of *C* in both treatments. A small fraction of subjects in *Cor-Chat* talk about non-lenient GRIM punishment, while subjects in *NoCor-Chat* do not attempt to agree on a punishment plan. Interestingly, although subjects play CGRIM, topics like the matching of the signals, or implicit and explicit punishment when signals differ, are rarely discussed. The frequency of their occurrence is below 0.1%.

To answer the question how communication is related to cooperation, we conduct logistic regressions. The marginal effects are presented in Table 6, with the dependent variable being the first-round cooperation, and the explanatory variables being the dummies whether or not the pre-play communication falls into a main category. We control for the supergame number and socio-demographic and other subject-related characteristics. Data include all supergames, because the cooperation in round one is so high in the last three supergames that little variation exists. Standard errors are bootstrapped with 1000 repetitions and are two-way clustered on subjects and matches (Cameron et al., 2011). In the treatment *Cor-Chat*, *Deliberation* is positively correlated to the first-round cooperation. In treatment *NoCor-Chat*, a positive correlation is found for the category *Trivia*. The effect of communication content on cooperation probably differs between treatments because subjects have more to learn when signals are correlated, making *Deliberation* important for sustaining cooperation.

We further assign subjects to strategies on the basis of posterior probabilities, that is we map each individual to a strategy that has the maximum posterior probability based on SFEM. Among the 136 subjects, 49 use correlation-based strategies (CGRIM or CTFT), 87 use non-correlation related strategies (ALLD, ALLC, TFT and WSLS). We look at the frequencies of main categorical communication regarding correlation-based strategies users and users of other strategies. Figure 5 presents the result.

[14]Detailed description of the sub-categories and the mapping from sub-categories to main categories are in Table G1 in the Supplementary Material. We use the same classification scheme as Dvorak and Fehrler (2024) except two differences. First, to account for signal correlation, we add sub-categories regarding matching of signals. Second, because communication in our sessions has a pre-play structure, we eliminate their category "Information" and its sub-categories which regard the share of information in repeated communication.

[15]When the rating is random, agreement occurs with probability $p_i^2(Yes) + p_i^2(No), i = 1, 2$, where $p_i(Yes)$ is the frequency of rater i classifying objects into any categories, and $p_i(No) = 1 - p_i(Yes)$.

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

Table 6. Communication and first round cooperation

	Cor-Chat	*NoCor-Chat*
Coordination	0.96	1.38
Deliberation	0.65**	0.09
Relationship	0.34	-0.21
Trivia	0.25	0.71*
Supergame	0.30***	0.11

Notes: The table presents the result of Logistic regression: the marginal effects of being in the main categories on first period cooperation. Data include all supergames. We control for socio-demographic and subject-related characteristics. Bootstrapped standard errors are calculated with two-way clustering, with cluster dimensions being subjects and match (1000 repetitions). Significance is based on the critical values of one-sided t statistics. * * * (**,*) indicates significance on the 1 (5,10)% level.

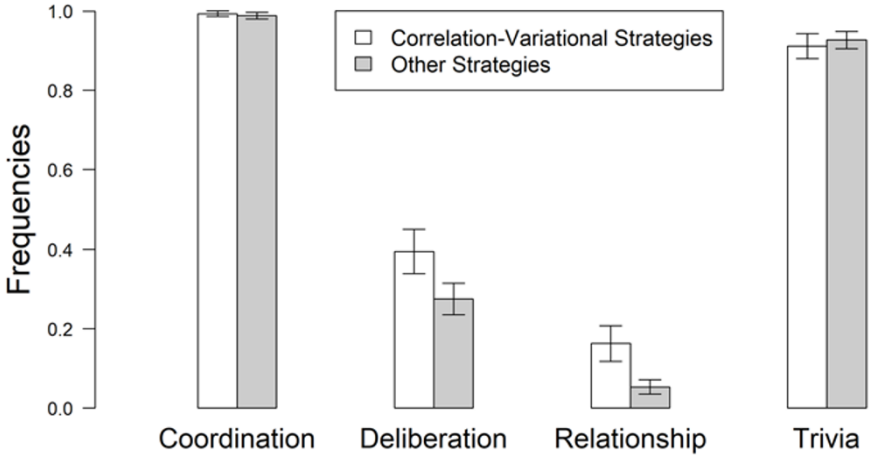


Figure 5. Frequency of communication of correlation-variants and other strategies
Notes: Bars show the frequencies of each communication category of subjects who use correlation-based strategies (CGRIM or CTFT) and other strategies. Individuals are assigned strategies with highest posterior probabilities. Data is from the last three supergames. Error bars represent two-way clustered standard errors at the subject and match levels.

Subjects who use correlation-related strategies have a similar communication pattern as the other subjects. There is only a small difference in *Deliberation* and *Relationship* ($p = 0.03$ and $p = 0.01$). Subjects who are assigned as users of CGRIM or CTFT deliberate more, and are involved in talks to build a relationship more often.

Subjects use pre-play communication to coordinate behavior. Subjects who engage in deliberation are more cooperative in the treatment with signal correlation, where deliberation is associated with the use of correlation-based strategies.

6. Conclusion

Sustaining cooperation is difficult in indefinitely repeated games when players cannot perfectly observe their opponents' actions. One way to sustain cooperation proposed in the theoretical literature is through correlated information (Awaya & Krishna, 2016; 2019). In a laboratory experiment, we investigate the effect of correlated information on strategy choice and cooperation in an indefinitely repeated Prisoner's Dilemma with public monitoring.

In the experiment, information is correlated in a simple way that is easy for subjects to understand. When players choose the same action, they receive identical signals about the actions of both players, which are the result of the same noisy process. If their actions are different, the two

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press

signals are independently determined by the same noisy process. With correlation, there is a simple grim-trigger strategy that can be used to efficiently support cooperation in equilibrium. This grim-trigger strategy cooperates as long as the two public signals exchanged in each round of the supergame match, and defects otherwise. Without correlation, such a strategy does not exist and cooperation cannot be supported in equilibrium. The theory therefore predicts more cooperation when signals are correlated. However, other studies have shown that people struggle to make correct inferences based on correlated information (e.g., Enke & Zimmermann, 2019; Fehrler et al., 2026). Therefore, whether the hypothesis that correlation can promote cooperation by improving the quality of imperfect monitoring would be confirmed was an open empirical question.

Under both *Chat* and *NoChat* conditions, we find no treatment difference in the average frequency of cooperation between treatments. Cooperation starts high in *Cor-Chat* and has a steeper decline. An analysis of participants' strategies shows that a significant fraction of participants in the *Cor-Chat* treatment take the correlation of signals into account when making their decisions. However, they do so by following a non-lenient punishment scheme for defection. Since not all participants in the *Cor-Chat* treatment are cooperative, punishments occur frequently. Similar patterns of strategy choice are observed in *Cor-NoChat*. This reveals two opposing roles of correlation. On the one hand, correlation makes it possible to detect defection with certainty, which makes efficient cooperation possible. On the other hand, defection becomes unambiguously detectable, and the response to defection is not lenient because there is no wiggle room around bad signals.

The discount factor we implemented in the lab is close to the threshold above which cooperative equilibria exist in the absence of correlation, which may explain why participants are also quite cooperative in the *NoCor* treatments. Social preferences, which have been shown to matter also in indefinitely repeated games (see, e.g., Kölle et al., 2025), offer a further explanation of the substantial cooperation rates in these treatments.

However, it is still puzzling why participants are so lenient and forgiving under *NoCor-Chat*. Dvorak and Fehrler (2024) observe high leniency also under imperfect private monitoring in combination with pre-play communication. This raises the broader question of why participants are lenient toward bad signals under imperfect monitoring when there is pre-play communication. One possible explanation is that pre-play communication allows participants to connect with each other, making social preferences more important (or easier to detect, as in Kartal & Müller, 2026). This makes subjects more lenient after noisy, bad signals, which need to be interpreted. Consistent with this idea, we first observe that subjects exhibit significantly lower levels of leniency and forgiveness in our *NoChat* treatments. Second, we observe a similar decrease in leniency between the *NoCor-Chat* and *Cor-Chat* treatments, in which monitoring becomes easier. Thus, communication combined with imperfect monitoring appears to be particularly conducive to cooperation. Future research could investigate this interaction in other settings.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/eec.2026.10069.

Replication package. The replication materials for this study are available at https://doi.org/10.6084/m9.figshare.33995040.

Acknowledgements. The study obtained ethical approval by the GfeW (No. C6C1g6Kj). A pre-analysis plan of the experiment was registered at the AEA registry as AEARCTR-0005369. AI tools (LLMs) have been used to polish some formulations in the manuscript. Fabian Dvorak gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984. Sebastian Fehrler gratefully acknowledges funding from the Juniorprofessorenprogramm Banden-Württemberg and from the Federal Ministry of Labour and Social Affairs (Germany) through the funding network Interdisciplinary Social Policy Research (FIS). ORCIDs: Yongping Bao, 0000-0001-6399-3300; Fabian Dvorak, 0009-0004-1109-5903; Sebastian Fehrler, 0000-0002-5220-5943.

References

Abreu, D., Pearce, D., & Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, *58*(5), 1041–1063.

Aoyagi, M., Bhaskar, V., & Fréchette, G. R. (2019). The impact of monitoring in infinitely repeated games: Perfect, public, and private. *American Economic Journal: Microeconomics*, *11*(1), 1–43.
Aoyagi, M., Fréchette, G. R., & Yuksel, S. (2024). Beliefs in repeated games: An experiment. *American Economic Review*, *114*(12), 3944–3975.
Awaya, Y., & Krishna, V. (2016). On communication and collusion. *American Economic Review*, *106*(2), 285–315.
Awaya, Y., & Krishna, V. (2019). Communication and cooperation in repeated games. *Theoretical Economics*, *14*(2), 513–553.
Backhaus, T., & Breitmoser, Y. (2026). Individual strategy choice in the repeated prisoner's dilemma. *European Economic Review*, *188*, 105407. https://doi.org/10.1016/j.euroecorev.2026.105407.
Bao, Y., Dvorak, F., & Fehrler, S. (2020). Exchanging private information to sustain cooperation in noisy, indefinitely repeated interactions. *AEA RCT Registry*, 5369–1.0. https://doi.org/10.1257/rct.5369-1.0.
Bland, J. R. (2025). Meta-study in the shadow of the future: The effect of continuation probability on strategies. Available at SSRN. https://doi.org/10.2139/ssrn.5042608.
Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, *71*, 117–120. https://doi.org/10.1016/j.euroecorev.2014.07.003.
Breitmoser, Y. (2015). Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma. *American Economic Review*, *105*(9), 2882–2910.
Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.
Compte, O., & Postlewaite, A. (2015). Plausible cooperation. *Games and Economic Behavior*, *91*, 45–59. https://doi.org/10.1016/j.geb.2015.03.010.
Cooper, R. W., DeJong, D. V., Forsythe, R., & Ross, T. W. (1992). Communication in coordination games. *Quarterly Journal of Economics*, *107*(2), 739–771.
Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, *101*(1), 411–429.
Dal Bó, P., & Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, *56*(1), 60–114.
Dal Bó, P., & Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma. *American Economic Review*, *109*(11), 3929–3952.
Dempster, A., Laird, N., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, *39*(1), 1–38.
Dvorak, F. (2022) *StratEst: Strategy Estimation*. R package version 1.1.6.
Dvorak, F., & Fehrler, S. (2024). Negotiating cooperation under uncertainty: Communication in noisy, indefinitely repeated interactions. *American Economic Journal: Microeconomics*, *16*(3), 232–258.
Ellingsen, T., & Östling, R. (2010). When does communication improve coordination? *American Economic Review*, *100*(4), 1695–1724.
Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, *86*(1), 313–332.
Fehrler, S., Hochleitner, A., & Janas, M. (2026). Too good to be true – Individual and collective decision-making with misleading signals. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2024.06504.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.
Fréchette, G. R., & Yuksel, S. (2017). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. *Experimental Economics*, *20*(2), 279–308.
Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, *102*(2), 720–749.
Green, E. J., & Porter, R. H. (1984). Noncooperative collusion under imperfect price information. *Econometrica*, *52*(1), 87–106.
Kartal, M., & Müller, W. (2026). A new approach to the analysis of cooperation under the shadow of the future: Theory and experimental evidence. *Quantitative Economics*, *17*(3), 827–857.
Kölle, F., Quercia, S., & Tripodi, E. (2025). Social preferences under the shadow of the future. *Experimental Economics*, *28*(1), 22–42.
Rabin, M. (1994). A model of pre-game communication. *Journal of Economic Theory*, *63*(2), 370–391.
Romero, J., & Rosokha, Y. (2023). Mixed strategies in the indefinitely repeated prisoner's dilemma. *Econometrica*, *91*(6), 2295–2331.
Sekiguchi, T. (1997). Efficiency in repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, *76*(2), 345–361.
Stigler, G. J. (1964). A theory of oligopoly. *Journal of Political Economy*, *72*(1), 44–61.

Cite this article: Bao, Y., Dvorak, F., & Fehrler, S. (2026). Sustaining cooperation with correlated information: an experimental test. *Experimental Economics*, 1–18. https://doi.org/10.1017/eec.2026.10069

https://doi.org/10.1017/eec.2026.10069 Published online by Cambridge University Press