Human and Algorithmic Cooperation: Parameters Matter![a]

Matthias Blonski Steffen Eibelshäuser Hendrik Hegemann
Victor Klockmann Alicia von Schenk[b]

November 2024

Abstract

Under which conditions do humans cooperate with machines? This study explores human and algorithmic cooperation in the infinitely repeated Prisoner's Dilemma, focusing on a critical threshold for cooperation. Using online experiments, we assess how replacing human opponents with learning algorithms – either human-mimicking or profit-maximizing – affects cooperation rates. Our results show that human cooperation is influenced by opponent type *and* strategy, with a higher predictive power of the theoretical threshold when interacting with AI maximizing profits.

JEL: C72, C73, C91, L13, O33

Keywords: Human-machine cooperation, Infinitely repeated Prisoner's Dilemma, Experiment, Learning algorithms, Collusion

---

[a]Financial support by the Hessian Center for Artificial Intelligence hessian.AI is gratefully acknowledged. We thank participants at the Thurgau Experimental Economics Meeting, the Workshop for Markets, Cooperation, and Votes in Madrid, the Workshop on Machine Behavior Research on Cooperative AI in Duisburg, and the behavioral and experimental groups in Würzburg and Frankfurt for valuable comments and suggestions.

[b]Blonski: Goethe University Frankfurt (email: blonski@wiwi.uni-frankfurt.de). Eibelshäuser: Goethe University Frankfurt (email: eibelshaeuser@econ.uni-frankfurt.de). Hegemann: Institute for Monetary and Financial Stability, Goethe University Frankfurt (email: hegemann@imfs-frankfurt.de). Klockmann: Julius-Maximilians-Universität Würzburg, Max Planck Institute for Human Development (email: victor.klockmann@uni-wuerzburg.de). von Schenk: Julius-Maximilians-Universität Würzburg, Max Planck Institute for Human Development (email: alicia.vonschenk@uni-wuerzburg.de).

1 Introduction

As machines become more sophisticated, particularly with the increasing deployment of *artificial intelligence* (AI) and advanced *learning algorithms* in economic contexts such as market pricing, understanding the conditions under which machines and humans cooperate effectively is of growing importance. A rapidly expanding body of literature is exploring the interaction and cooperation between humans and various types of learning algorithms, as well as cooperation among algorithms themselves. At the same time, a debate has emerged about whether the replacement of humans by machines in oligopolistic markets could encourage collusion and thereby reduce competition. This debate is not yet settled. For example, the German Monopolkommission, a publicly funded think tank advising policymakers, expressed in its main report of 2018 that there are many obstacles to coordination and communication among machines and humans. Similarly, researchers such as Kühn and Tadelis (2018) and Schwalbe (2018) are among the less concerned, demonstrating that it remains difficult to build up collusion in most real world environments. However, others such as or Crandall et al. (2018), Calvano et al. (2020), Klein (2021) or Normann and Sternberg (2023) have shown in their more specific contexts that increasingly sophisticated machines indeed can learn to cooperate in experimental settings.

Our current project aims to contribute to this debate by focusing on the leading question: *Under which conditions do machines and humans cooperate with each other?* We address this question within the framework of the two-player infinitely repeated discounted Prisoner's Dilemma as a simple and fundamental setting that is omnipresent in the social sciences and beyond. This setting is well-suited for our study since it has been the subject of extensive research, allowing us to compare our novel observations with decades of experimental evidence of human-human interactions (see, e.g., literature reviews by Dal Bó and Fréchette; 2018; Mengel; 2018). Additionally, for the repeated discounted Prisoner's Dilemma paradigm, we can build on recent theories which proposed an equilibrium selection criterion based on strategic risk that has been validated in experimental labs (Blonski et al.; 2011; Blonski and Spagnolo; 2015). This δ^*-*criterion*, as we will refer to it, predicts a substantial increase of cooperation rates among humans when the continuation probability δ of the game surpasses a critical threshold δ^*, which depends on all exogenous payoff parameters of the Prisoner's Dilemma. In contrast to many other projects, we explicitly allow *interaction between humans* *and machines*, rather than limiting our analysis to human-human or machine-machine interactions. Our treatment variations focus on opponent type (human vs. machine) and behavior (objective and actions), which is central to our investigation of how machine opponents affect human cooperation. In one group, subjects play against other humans (Human treatments). In another, they engage with a human-behavior-mimicking algorithm (Imitator treatments), trained using data from the Human treatments. A third group faces an algorithm designed to maximize its own payoffs (Optimizer treatments). These three treatments help us disentangle the different factors influencing cooperation behavior: the opponent's type (human vs. machine) and the behavior of the algorithm (human-like vs. profit-maximizing). Determining whether changes in cooperation behavior among subjects are due to the altered type of opponent or the change in the opponent's behavior sets our study apart from previous experiments where humans play against computer players (see literature reviews by March; 2021; Chugunova and Sele; 2022).

What do we find? We considered three different continuation probabilities, which predict minimal and maximal cooperation levels, based on equilibrium selection theory. The first set of results focuses on human cooperation across opponent types. As expected, we observed a clear increase in cooperation rates from the low to the medium to the high continuation probability. Yet, a significant proportion of human participants still turned out to cooperate with each other under the medium continuation probability and even under the low continuation probability.[1] When human opponents are replaced by two types of algorithms – a human-imitator algorithm and a profit-maximizing algorithm – a distinct three-stage pattern emerges across all three continuation probabilities. Regardless of the continuation probability, participants exhibit statistically significantly higher levels of cooperation when faced with human opponents compared to their algorithmic counterparts that mimic human behavior. Furthermore, our evidence reveals even lower cooperation rates in Optimizer treatments. The second of results concerns the δ^*-selection criterion, which predicts a discontinuous increase

[1]These cooperation rates for the low and medium δ were higher compared to various former experiments, including the original experiments reported in Blonski et al. (2011). Possible explanations for this could be that previous experiments were conducted in university labs where almost all participants are students and experimenters have control over subjects' information and communication. Our experiments, however, were conducted on the Prolific platform, where the participant pool likely receives less formal training compared to university students but may be more representative of society at large. In addition, available information and communication is more messy but may be more realistic.

in cooperation rates only once the continuation probability δ surpasses the threshold δ^*. We propose a two-dimensional measure to assess its predictability. First, the criterion suggests that the difference in initial cooperation remains unchanged when the parameters remain on the same side of the δ^*-threshold. Second, it predicts that the difference in initial cooperation is maximal once the parameters exceed the δ^*-threshold. It turns out that in both dimensions, predictability of equilibrium selection theory increases if humans are replaced by profit maximizing algorithms. It is less surprising that human players who face an opponent that consistently initiates cooperation or defection at some point recognize that mirroring this behavior benefits them. However, what we did not expect was a 100%-alignment for initial cooperation of the profit-maximizing algorithm with the theoretical threshold. This reinforcement learning algorithm was trained against the human imitator algorithm that behaved just as erratically as the human participants. Nevertheless it quickly learned that the most effective way to maximize profits was to initiate cooperation precisely as predicted by the δ^*-criterion. This result is in itself interesting, as neither the reinforcement learner nor the human participants were informed about equilibrium selection theory.

To summarize, our findings offer new insights into the conditions under which machines and humans cooperate in the infinitely repeated Prisoner's Dilemma. Specifically, we find that human cooperation rates react to both, the nature of the opponent (human vs. machine) and its objective (imitate humans or maximize profits). Further, the predictive power of the δ^*-criterion, based on strategic risk, becomes *stronger* when a human plays against a machine instead of another human. It becomes *particularly strong* if the algorithm is designed to learn how to maximize profits against humans, as many real-world algorithms are. The prediction is nearly 100% accurate when two profit-maximizing algorithms play against each other in simulations. Profit-maximizing algorithms figure out the same cooperation criterion as abstract game theory does, but through an entirely different approach: processing data from human populations. One way to interpret these observations is that the introduction of algorithms into the erratic realm of humans increases predictability for cooperation and collusion. Notably, this does not imply that the introduction of algorithms generally decreases or increases cooperation. It depends strongly on the degree of algorithmic integration and the parameters of the strategic setting.

2 Experimental Design

To investigate the impact of machine opponents on human cooperative behavior, we conducted a between-subjects empirical study using an online experiment based on the infinitely repeated discounted Prisoner's Dilemma. The experiment was pre-registered prior to execution.[2] In implementing the experiment on an infinitely repeated discounted Prisoner's Dilemma, we built on the standard approach proposed by Roth and Murnighan (1978), wherein the game terminates with a probability of $1 - \delta$ after each stage game. For a risk neutral player the continuation probability δ is equivalent to the discount factor of the player. Each subject engages in a total of 10 runs of the repeated game ("supergames"), with a stranger matching between two subsequent supergames.

2.1 Opponent Type and Behavior

Our main treatment variation, namely opponent type or nature – i.e., human vs. machine – and behavior – i.e., objective and chosen actions – is central to our key question regarding the effects of machine opponents on human cooperation behavior and sets us apart from previously conducted experiments involving infinitely repeated Prisoner's Dilemmas. In our between-subjects experimental design, subjects were assigned to one of three distinct treatment groups, each differing in the type and behavior of their opponent. One group of subjects plays against other subjects within the group (Human treatments). In another group, subjects engage with a human behavior-mimicking algorithm (Imitator treatments), trained using data from the Human treatments. In a third group, subjects face off against an algorithm designed to maximize its own supergame payoff. This algorithm has been trained to play against the human-behavior-mimicking algorithm (Optimizer treatments). Participants were explicitly informed about the type of opponent they would face prior to the start of the experiment. This included whether the opponent was another human, a human-behavior-mimicking algorithm, or a profit-maximizing algorithm. After each stage game, participants received feedback on their opponent's decisions and rewards from the ongoing game.

These three treatments allow us to identify the different channels that influence subjects' cooperation behavior. In the Imitator treatments, the type of opponent player differs from

[2] AsPredicted #130373, https://aspredicted.org/9WB_YZR

that in the Human treatments, while the behavior of the (now machine) opponent is designed to match human play as closely as possible. In the Optimizer treatments, the opponent player's type remains the same as in the Imitator treatments (both are algorithms), but the algorithm's behavior shifts to profit maximization. Through these three treatments, we can ascertain whether altered cooperation behavior among subjects is attributable to the changed type of the opponent player or the change in the opponent player's behavior.[3]

The two types of algorithmic players, the human imitator and the self-interested optimizer, are implemented generically, aiming for the fewest model assumptions possible. The human behavior-mimicking algorithm is model-free and replicates the actions played in the Human treatment. At the beginning of a supergame, a human player from the Human treatment (with the same continuation probability) is drawn at random, and her actions in the experiment are used to sample actions of the imitator for the entire supergame. In this way, individual idiosyncrasies of human players can be reflected in the imitator algorithm. In every stage game of the supergame, only actions of the human player in comparable situations are considered. If the human has faced the same history of actions in the supergame up to this stage game as the imitator, only the actions played by the human after this history are used to sample the next action of the imitator. If the human has not faced the same history, only a partial history is considered. In the rare case of no overlap in action histories of human player and imitator, action data from other humans of the Human treatment is used to sample the next action of the imitator.

The self-interested optimizer is implemented as a standard reinforcement learning algorithm (Sutton; 1988; Rummery and Niranjan; 1994; Sutton and Barto; 1998). In contrast to the imitator, the reinforcement learner must be based on a model so that its parameters can be determined during the learning process. To avoid imposing assumptions and restrictions, we have considered a vast set of possible models and have chosen the best model for our use case – in the sense that the resulting algorithm's objective, namely maximizing its own payoff against human imitators, is realized best. The final model features 17 states, enabling the algorithm to base its actions on the most relevant parts of the history of play in the

[3]The money earned by the algorithmic players was not paid out but remained in the experimenters' budget. The instructions did not mention what happened with the algorithm's earnings, which participants usually understand as the money not being paid to any human entity (von Schenk et al.; 2023).

supergame. Specifically, it considers the actions taken by both players in the previous period (4 possibilities), the actions in the very first period (4 possibilities), and includes a distinct state for the initial period of the supergame. Our result that the previous as well as the initial period matter the most is consistent with the literature on the Repeated Prisoner's Dilemma, see for example Breitmoser (2015), Dal Bó and Fréchette (2019) and Fudenberg and Karreskog Rehbinder (2024). Detailed technical specifications of the algorithm in Imitator and Optimizer treatments can be found in Appendix A.

2.2 Parametrization of the Repeated Prisoner's Dilemma

In Blonski and Spagnolo (2015) and Blonski et al. (2011), it has been demonstrated both theoretically and experimentally that there exists a critical parameter range for the repeated discounted Prisoner's Dilemma in which the cooperation rates of players change significantly. In our exploration of cooperation behaviors influenced by human and machine opponents, we build upon their equilibrium selection criterion, specifically the δ^*-criterion, when choosing our parametrization. In the following, we briefly revisit the corresponding equilibrium selection theory.

Blonski and Spagnolo (2015) and Blonski et al. (2011) build upon the work of Harsanyi and Selten (1988), hereafter referred to as HS. Blonski and Spagnolo (2015) showed that the risk dominance concept of HS for simple $2 \times 2-$games is applicable in a consistent manner to the infinitely repeated discounted Prisoner's Dilemma. They formulated an equilibrium selection criterion based on HS's $2 \times 2-$games risk dominance. Blonski et al. (2011) formulated axioms building on HS's axiomatic foundation of risk dominance for $2 \times 2-$games and adapted them to the dynamic nature of the infinitely repeated discounted Prisoner's Dilemma. Both approaches identify the same δ^*-criterion. By conducting experiments across the relevant parameter range, Blonski et al. (2011) demonstrated that the δ^*-criterion predicts cooperation of players in the lab with high accuracy. This criterion performs particularly well compared to the classical criterion $\underline{\delta}$, which assumes that players will always cooperate once the cooperation is supported by an equilibrium.

To be more precise, consider the infinitely repeated discounted repetition of the normalized, symmetric Prisoner's Dilemma stage game characterized only by the three parameters g, l, δ, interpreted as gain $g > 0$, loss $l > 0$ and continuation probability or discount factor $\delta \in (0, 1)$:

	cooperate	defect
cooperate	$1\ ,\ 1$	$-l\ ,\ 1+g$
defect	$1+g\ ,\ -l$	$0\ ,\ 0$

The inequality

$$\delta \geq \underline{\delta} = \frac{g}{g+1}.$$

defines the lower bound on players' patience to support equilibria with cooperative actions on its outcome path. In many applications, the size of the interval $[\underline{\delta}, 1]$ with its lower bound $\underline{\delta} = \frac{g}{g+1}$ has been interpreted and is still used as a criterion for how well players can cooperate. This is why we call $\underline{\delta}$ the "classical" criterion. One possible justification for this interpretation is to consider Pareto dominance as the most relevant equilibrium selection criterion, instead of risk dominance. When both cooperative and non-cooperative equilibria are present, the indefinitely defective outcome is Pareto-dominated by the indefinitely cooperative path, given $\frac{1}{1-\delta} > 0$. However, an obvious deficiency of the classical criterion $\underline{\delta} = \frac{g}{g+1}$ is that it does not depend on the loss l suffered by the player who cooperates alone, usually referred to as *sucker's payoff*. Hence, using the classical criterion $\underline{\delta}$, the strategic risk of picking any cooperative equilibrium can get arbitrarily large for sufficiently large loss l.

The δ^*-criterion

$$\delta \geq \delta^* = \frac{g+l}{1+g+l} > \underline{\delta} = \frac{g}{g+1}$$

predicts cooperation if $\delta > \delta^*$ and addresses the latter problem. There are two different theoretical foundations for it, both of which refer to strategic risk which in this parsimonious 3-parameter version of the model can simply be quantified by the loss parameter l. Equally important in our context is that $\delta^* = \frac{g+l}{1+g+l}$ identifies the critical parameter range where a significant jump in cooperation rates is expected. If equilibrium selection based on strategic risk has any relevance for real-world players' tendency to cooperate, they should cooperate much more once $\delta \geq \delta^*$. The experimental evidence in Blonski et al. (2011) clearly falsified the classical criterion $\underline{\delta}$ and its comparative statics, which has been used for decades and still persists in the applied community. Beyond that, it supported the δ^*-criterion. Moreover, it was also consistent with variations in cooperation rates observed in previous studies and led to follow-up projects such as Breitmoser (2015) and Fudenberg and Karreskog Rehbinder (2024).

For each of the three treatments with a different opponent (Human, Imitator and Optimizer), we selected different continuation probabilities δ that lie both above and below the δ^*-criterion. The payoffs for the stage game in our experiment remain constant and consistent across all treatments, as detailed in Figure 1.

		Player 2	
		C	D
Player 1	C	$90, 90$	$0, 120$
	D	$120, 0$	$40, 40$

		Player 2	
		C	D
Player 1	C	$1, 1$	$-4/5, 8/5$
	D	$8/5, -4/5$	$0, 0$

Figure 1: Stage Game Payoffs

Notes: The left figure displays the payoffs of the stage game in Experimental Currency Units (ECU), where £1=1000 ECU. The right figure displays the payoffs in normalized form.

The payoffs outlined in Figure 1 yield $\delta^* = 7/12 \approx 0.583$. The "classical" criterion is $\underline{\delta} = 0.375$ for these payoffs. We considered three different continuation probabilities: $\delta_1 = 0.85$, $\delta_2 = 0.55$, and $\delta_3 = 0.4$. In treatments with δ_1, cooperation is supported as an equilibrium by δ^*-selection criterion, whereas for δ_2 and δ_3, it is not. The selection of these three different δ values encompasses the entire spectrum for which δ^* makes predictions, namely $\delta \in [\underline{\delta}, 1]$, and encompasses all critical values within this spectrum. Specifically, δ_1 significantly exceeds the critical value of the δ^*-selection criterion. In contrast, δ_2 is close to the critical value δ^*, while δ_3 is close to the "classical" criterion, $\underline{\delta}$. Table B.1 in the Appendix provides an overview of the different treatments and the number of participants.[4]

[4]Contrary to our pre-registration, we implemented only one instead of two δ values above δ^* in the experiment. Nevertheless, the chosen δ values still cover a substantial portion of the relevant spectrum for which δ^* makes predictions. Moreover, the number of independent observations per δ and per type of the opponent, approximately 150 (i.e., 300 observations for each Human treatment), aligns with the planned number in the pre-registration.

2.3 Experimental Procedures

After receiving instructions for the experiment, participants were tasked with correctly answering a series of comprehension questions related to stage game payoffs and continuation probability. Individuals who failed to answer these questions correctly on three separate occasions were excluded from further participation in the experiment. Upon completion of all supergames, the subjects' beliefs regarding the type and strategy of their opponents were obtained through an incentivized elicitation.

The experiment was conducted online via Prolific with 1,847 participants between September 2023 and January 2024. For implementation, oTree by Chen et al. (2016) was utilized and hosted on Heroku servers. Subjects took between 15 and 25 minutes to complete the experiment, depending on the specific δ and opponent type of the treatment. On average, subjects received payments totaling £4.70, inclusive of a base payment of £2.50. Instructions and comprehension questions for the experiment can be found in Appendix D.

3 Results

Our main focus lies on the average cooperation rates in the first round of each supergame. Since we did not include trial rounds after the instructions, we exclude the first three of the ten supergames each participant played if not mentioned otherwise (see Blonski et al. (2011) who also proceeded in that way). Besides initial cooperation rates, we also report overall cooperation rates in Table B.2 in the appendix, which show qualitatively the same pattern.

3.1 Cooperation Across Opponent Type

When comparing the rates of initial human cooperation between different opponent types (see Figure 2), a distinct three-stage pattern across all three δ values emerges. Regardless of the continuation probability, participants exhibit statistically significantly higher levels of cooperation when facing human opponents compared to algorithmic counterparts imitating human behavior ($p < 0.001$ for $\delta \in \{0.4, 0.55\}$, and $p = 0.004$ for $\delta = 0.85$, Wald tests). Additionally, our analysis reveals even lower cooperation rates in the Optimizer treatment for continuation probabilities below δ^* (all $p < 0.001$, Wald tests).

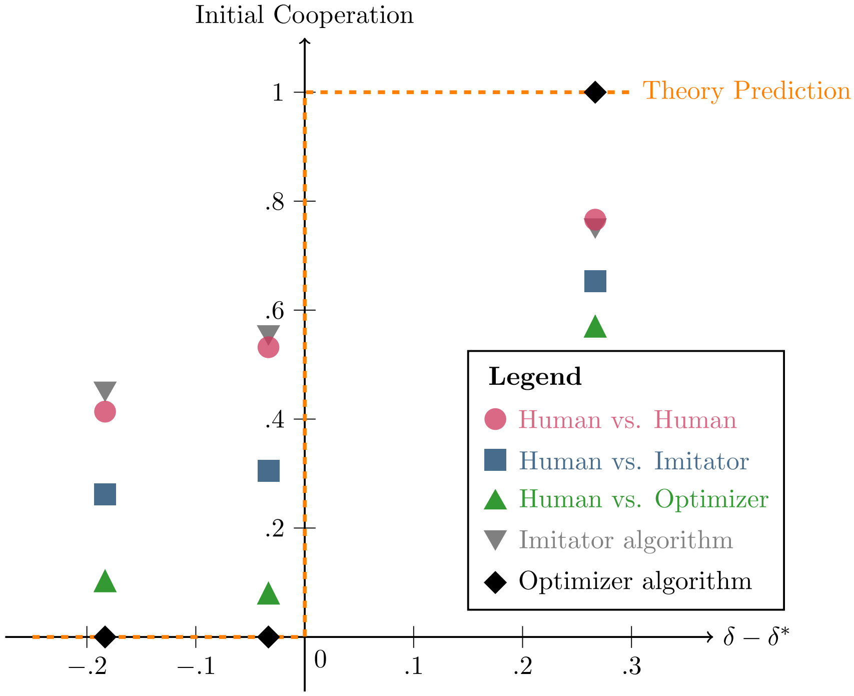

Figure 2: Initial Cooperation Rates

Notes: The figure displays the initial cooperation rates by continuation probability and by treatments varying opponent type and behavior.

Result 1 (Initial cooperation across opponent types) *For all δ, participants cooperate significantly less as algorithmic integration rises.*

To our knowledge, our study stands out for its stringent differentiation between variations in willingness to cooperate in human-machine interactions stemming from differences in the machine's choices versus differences in the machine's inherent nature. In summary, participants cooperate significantly more with other humans than with an algorithm imitating human choices (for all δ). Cooperation decreases further when they interact with a payoff-maximizing reinforcement learning algorithm (if $\delta < \delta^*$). Thus, our findings suggest that both the type and strategy of the opponent play crucial roles in shaping initial cooperation rates within the repeated Prisoner's Dilemma framework.

In the treatments with algorithmic opponents, we can observe the initial cooperation rates of the machine players, as depicted in Figure 2 and detailed in Appendix A. The human imitator plays a completely mixed strategy. Its probabilities of cooperation feature considerable heterogeneity across supergames, reflecting the heterogeneity in cooperation across human players. Both on average and on the individual level, the cooperation probabilities of the imitator closely follow those of the participants in the Human treatments – by design (minor deviations due to randomness in simulations). In contrast, the reinforcement learner adopts a corner solution, adhering precisely to the the δ^*-criterion by never cooperating for the two lower δ values and consistently cooperating for the high continuation probability of 0.85. This behavior is intriguing, particularly considering that the reinforcement learner was trained to optimize its own payoffs against our participant sample (represented by the human imitator algorithm) which exhibits overall high cooperation rates. Nevertheless, it is noteworthy that the reinforcement learner consistently refrains from cooperation when $\delta < \delta^*$, even if the opponent chooses to cooperate.

Result 2 (Algorithmic cooperation) *A reinforcement learner algorithm that seeks to maximize its own payoffs plays the corner solution according to δ^*-selection criterion and cooperates if and only if $\delta > \delta^*$.*

Despite the full cooperation of the reinforcement learning algorithm for the high continuation probability $\delta = 0.85$, human participants cooperate significantly less against this type of opponent than when playing against another human (cf. Result 1). There are two opposing mechanisms in this experimental condition: the cooperative strategy of the opponent, which should encourage cooperation by the participants, and its algorithmic nature, which discourages cooperation, as it can be seen in the comparison between the Human and Imitator treatments. The low human cooperation rate indicates that the latter mechanism exceeds the former.

In the experiment, each participant engages in a total of 10 supergames, with a stranger matching implemented between two subsequent supergames. Apart from examining the aggregated initial cooperation rates, we proceed to analyze the evolution of these rates across supergames for human participants (see Figure 3). Although initial cooperation rates in the first supergame exceed 50% across all opponent types, sustained high cooperation rates are only observed for the highest continuation probability, $\delta = 0.85$. Conversely, for the two δ values below δ^*, initial cooperation rates gradually decline over time irrespective of the opponent, indicating a convergence towards the δ^*-criterion, particularly noticeable in the algorithm treatments after numerous supergames. This decline is most pronounced when

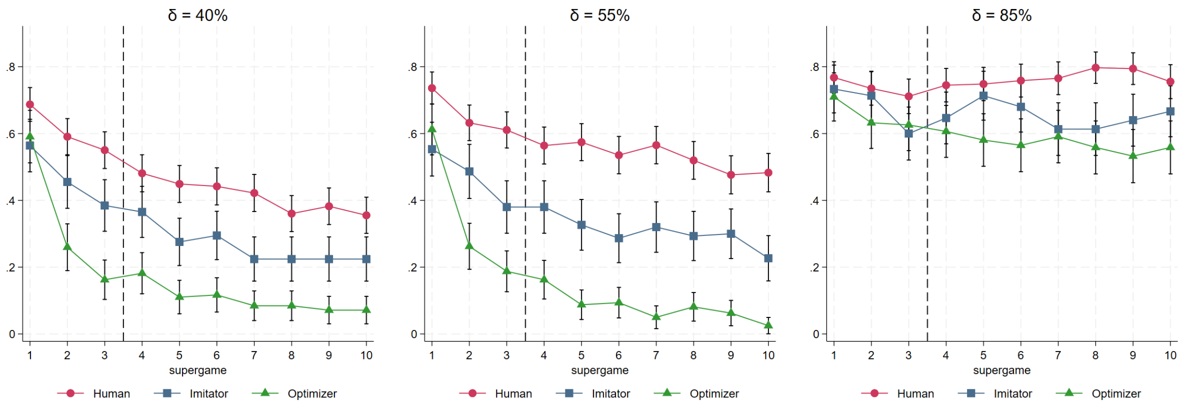

Figure 3: Evolution of Initial Cooperation

Notes: The figure displays the evolution of participants' initial cooperation rates for all types of the opponent across supergames, separately by the continuation probability δ. The dashed vertical line separates the excluded first three periods from the main dataset used for the analyses.

participants play against the optimized reinforcement learning algorithm, where cooperation for $\delta < \delta^*$ nearly diminishes to zero by the 10th supergame. As reported in Result 2, the reinforcement learning algorithm never cooperates for $\delta < \delta^*$. Consequently, human participants show a sharp decrease in cooperation between the first and the second supergame. Afterwards, though they require some time, they eventually converge to the algorithm's behavior in the long run in the Optimizer treatment. The visual impression is supported by regression results reported in Table B.3 in the appendix. The initial differences between the three opponent types further increase by time trends that are more negative in the Imitator treatment ($p = 0.06$) and even stronger in the Optimizer treatment ($p < 0.005$).

Result 3 (Evolution of human cooperation) *The gap in cooperation rates across opponents widens over time. That is, participants need several supergames to adjust their strategy to the type of opponent.*

3.2 Predictability of the δ^*-Selection Criterion

The δ^*-selection criterion predicts a discontinuous increase in cooperation rates only once the continuation probability δ surpasses the threshold δ^*. To measure the fit of observed human behavior for the different types of opponent, we define our measure of δ^*-*predictability* in two dimensions: first, the difference in initial cooperation if $\delta < \delta^*$ (in our case $\delta \in \{0.4, 0.55\}$),

	(1) Human	(2) Imitator	(3) Optimizer	(4) All
$\mathbb{1}\{\delta = 0.55\}$	0.118***	0.0429	-0.0226	0.118***
	(0.0329)	(0.0369)	(0.0181)	(0.0329)
Imitator				-0.152***
				(0.0329)
Optimizer				-0.311***
				(0.0265)
Imit $\times$ $\mathbb{1}\{\delta = 0.55\}$				-0.0753
				(0.0494)
Opt $\times$ $\mathbb{1}\{\delta = 0.55\}$				-0.141***
				(0.0375)
Constant	0.414***	0.262***	0.103***	0.414***
	(0.0231)	(0.0234)	(0.0131)	(0.0231)
N	4300	2142	2198	8640
R^2	0.014	0.002	0.002	0.122

(a) Comparisons $\delta = 0.40$ vs. $\delta = 0.55$.

	(1) Human	(2) Imitator	(3) Optimizer	(4) All
$\mathbb{1}\{\delta > \delta^*\}$	0.294***	0.370***	0.479***	0.294***
	(0.0273)	(0.0379)	(0.0368)	(0.0273)
Imitator				-0.189***
				(0.0248)
Optimizer				-0.381***
				(0.0189)
Imit $\times$ $\mathbb{1}\{\delta > \delta^*\}$				0.0766
				(0.0467)
Opt $\times$ $\mathbb{1}\{\delta > \delta^*\}$				0.185***
				(0.0458)
Constant	0.472***	0.283***	0.0914***	0.472***
	(0.0166)	(0.0184)	(0.00905)	(0.0166)
N	6318	3192	3278	12788
R^2	0.076	0.126	0.271	0.192

(b) Comparisons $\delta < \delta^*$ vs. $\delta > \delta^*$.

Table 1: δ^*−Predictability – Regressions

Notes: Panel (a) shows regression results of initial cooperation on an indicator for $\delta = 0.55$, treatment dummies, and interaction terms, and includes only observations with $\delta = 0.4, 0.55 < \delta^*$. Panel (b) shows regression results of initial cooperation on an indicator for $\delta > \delta^*$, treatment dummies, and interaction terms, and includes observations for all levels of δ. Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

which is predicted to be 0; second, the difference in initial cooperation between $\delta = 0.85 > \delta^*$ and $\delta = 0.4, 0.55 < \delta^*$, which is predicted to equal 1.

When participants engage with a human opponent, we observe a significant increase in cooperation rates when δ exceeds δ^* (see Figure 2 and column (1) of Table 1, panel (b)). Still, we find deviations from behavior anticipated by the δ^*−selection criterion. Predictions based on this criterion suggest negligible and indistinguishable cooperation for $\delta \in \{0.4, 0.55\}$, given their position below $\delta^* = 0.5833$. However, our observations reveal markedly more cooperation than predicted by the δ^*−selection criterion for these lower continuation probabilities (47.2% on average, see Table 1, panel (b)), and significant differences between $\delta = 0.4$ and $\delta = 0.55$ ($p < 0.001$, see Table 1, panel (a)).

Yet, when participants engage with algorithmic opponents, we find closer alignment with the δ^*−selection criterion. For the Imitator treatment, as reported in column (4) of Table 1, panel (b), cooperation rates for $\delta < \delta^*$ are significantly closer to zero ($p < 0.001$). Further, the difference in cooperation for $\delta > \delta^*$ vs. $\delta < \delta^*$ is larger compared to the Human treatment, though not significantly so ($p = 0.101$), and the difference for $\delta = 0.4$ vs. $\delta = 0.55$ mostly disappears and becomes insignificant ($p = 0.247$, Table 1, panel (a)). The largest effect of surpassing δ^* can be observed in the Optimizer treatment. When participants play against an algorithm that seeks to maximize its profits, they show on average an initial cooperation rate below 10% whenever $\delta < \delta^*$. The second dimension of δ^*-predictability – the increase in cooperation once $\delta > \delta^*$ – is significantly higher in the Optimizer treatment (0.479) than in the Human treatment (difference of 0.185, $p < 0.001$) and than in the Imitator treatment (difference of 0.109, $p = 0.040$, Wald test). Beyond that, as in the Imitator treatment, there is no significant difference in cooperation between $\delta = 0.4$ and $\delta = 0.55$ ($p = 0.212$).

Consequently, we observe that human behavior increasingly aligns with the δ^*-equilibrium selection criterion as the degree of algorithmic integration rises. When playing against another human, participants show a high willingness to cooperate, even for continuation probabilities $\delta < \delta^*$, and δ^*-predictability is rather poor with significant differences in cooperation below δ^* (0.118) and only a small increase when surpassing δ^* (0.294). However, if the opponent is replaced by an algorithm that imitates human choices, cooperation rates start to align more closely with the predictions of the δ^*-criterion, though the improvements in δ^*-predictability (to 0.0429 in the first dimension and to 0.370 in the second dimension) just miss significance ($p = 0.101$ and $p = 0.128$, respectively) and we still observe initial cooperation in 28.3% of cases for $\delta < \delta^*$. Finally, when participants engage with an algorithm trained to maximize payoffs using self-learned strategies, their behavior shifts toward patterns of initial cooperation similar to those predicted by the δ^*-selection criterion. In this scenario, cooperation rates are low for $\delta < \delta^*$ (around 10%), and δ^*-predictability is significantly stronger compared to human-human interactions (improvements to -0.0226 and to 0.479 with $p < 0.001$ in both dimensions).

Result 4 (δ^*-Predictability) *Human behavior aligns significantly more strongly with the δ^*-equilibrium selection criterion when interacting with a payoff-maximizing reinforcement learning algorithm than when interacting with another human.*

The nature of an experiment on repeated games does not allow for the observation of strategies, but only of outcome paths, which can contain equilibrium or off-equilibrium behavior.

		δ		
		0.4	0.55	0.85
Human	Payoff human	57.20	61.37	69.57
Imitator	Payoff human	59.89	64.52	66.31
	Payoff algorithm	46.18	47.07	63.39
Optimizer	Payoff human	37.09	37.40	87.37
	Payoff algorithm	47.19	45.20	56.12

Table 2: Average Payoffs

Notes: The table displays the average payoffs of humans and bots in all treatments and for all levels of δ.

As a robustness check on Results 1 and 4, we employ the two-step filtering procedure in Blonski et al. (2011) to estimate initial cooperation in the subset of equilibrium outcome paths in the data. Results are reported in Appendix C. The patterns of initial cooperation across levels of δ and opponent type are qualitatively unaffected. In the Optimizer treatment, the participants' behavior gets even closer to the prediction by the δ^*-criterion, with negligible cooperation in the first period below 4% whenever $\delta < \delta^*$.

3.3 Payoffs of Human and Algorithmic Players

As an exploratory, non-preregistered analysis, we consider the payoffs of human and algorithmic players and discuss under which conditions delegating decision authority to an algorithm could be desirable in order to maximize payoffs. Table 2 reports average payoffs of human and algorithmic players across all treatments and continuation probabilities. For $\delta < \delta^*$, participants earned significantly less payoff against an optimizer algorithm (which acts fully defectively, see Result 2) compared to the case of a human or imitator opponent ($p < 0.001$, Wald tests). For $\delta > \delta^*$, however, human subjects profit from the cooperative nature of the optimizer algorithm and achieve the highest payoffs across all conditions. Comparing the outcome of human and algorithmic players in the Optimizer treatments, the latter achieve higher payoffs for $\delta < \delta^*$ by exploiting the cooperation intentions of their human opponents ($p < 0.001$, paired t-tests). Under the high continuation probability, the pattern switches and the human players outperform their algorithmic opponent in terms of payoffs ($p < 0.001$).

Result 5 (Payoffs) *When playing against a reinforcement learner, delegating decisions to this algorithm on average maximizes players' payoffs if $\delta < \delta^*$; otherwise, making decisions oneself is payoff-maximizing. When playing against humans, it is always payoff-maximizing not to delegate and to decide on one's own.*

Imagine a population of human players, who can pick actions by themselves or delegate their actions to a machine of their choice. Suppose these players are matched randomly to play the infinitely repeated PD games under consideration. Result 5 suggests that a monotone dynamic in this evolutionary-game-style thought experiment tends to converge to either extreme where either (i) *only humans or human-like algorithms*, or (ii) *only profit-maximizing algorithms* prevail. It depends very much on parameters, however, as the title of this article emphasizes, which of the two cases realizes.

4 Conclusion

Our study provides a nuanced perspective on the conditions under which machines and humans cooperate in the infinitely repeated Prisoner's Dilemma, a foundational setting for studying cooperation and collusion. By systematically varying opponent type (human vs. machine) and algorithmic behavior (human-imitating vs. profit-maximizing), we demonstrate that human cooperation rates depend significantly on both the nature and objectives of their counterparts.

The introduction of equilibrium selection theory and strategic risk into infinitely repeated games with its plethora of equilibria has enhanced the predictability of human cooperation. Our current findings further support the potential usefulness of this decades-old theoretical paradigm. A key insight is that the predictive power of equilibrium selection theory increases when humans face machines, particularly when the machine's objective is profit maximization. When humans interact with these algorithms, they too become more predictable, provided they can learn from the algorithms' superior performance and emulate them in this context. We conjecture that human attitudes toward technology will change over time and across cultures. However, we expect the influence of introducing algorithms on the predictability of cooperation to be more likely to persist.

These findings contribute to the ongoing debate about algorithmic cooperation and competition. While our results suggest that algorithms may enhance predictability in strategic interactions, they also underscore the complexity of evaluating their broader implications. The degree of algorithmic integration and the specific parameters of the strategic environment play a crucial role in determining whether cooperation is increased, reduced, or merely transformed. Future research can build on this replicable experimental framework to explore more complex settings, such as with more than two players, larger action spaces, and incomplete information, further advancing our understanding of human-machine cooperation in economic and social contexts.

References

Blonski, M., Ockenfels, P. and Spagnolo, G. (2011). Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence, *American Economic Journal: Microeconomics* **3**(3): 164–192.

Blonski, M. and Spagnolo, G. (2015). Prisoners' other dilemma, *International Journal of Game Theory* **44**: 61–81.

Breitmoser, Y. (2015). Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma, *American Economic Review* **105**(9): 2882–2910.

Calvano, E., Calzolari, G., Denicolo, V. and Pastorello, S. (2020). Artificial intelligence, algorithmic pricing, and collusion, *American Economic Review* **110**(10): 3267–3297.

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

Chugunova, M. and Sele, D. (2022). We and it: An interdisciplinary review of the experimental evidence on how humans interact with machines, *Journal of Behavioral and Experimental Economics* **99**: 101897.

Crandall, J. W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J.-F., Cebrian, M., Shariff, A., Goodrich, M. A. and Rahwan, I. (2018). Cooperating with machines, *Nature Communications* **9**(1): 233.

Dal Bó, P. and 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. and Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma, *American Economic Review* **109**(11): 3929–3952.

Fudenberg, D. and Karreskog Rehbinder, G. (2024). Predicting cooperation with learning models, *American Economic Journal: Microeconomics* **16**(1): 1–32.

Harsanyi, J. C. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, MA.

Klein, T. (2021). Autonomous algorithmic collusion: Q-learning under sequential pricing, *The RAND Journal of Economics* **52**(3): 538–558.

Kühn, K.-U. and Tadelis, S. (2018). The economics of algorithmic pricing: Is collusion really inevitable?, *Working paper*, University of California, Berkeley.

March, C. (2021). Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players, *Journal of Economic Psychology* **87**: 102426.

Mengel, F. (2018). Risk and temptation: A meta-study on prisoner's dilemma games, *The Economic Journal* **128**(616): 3182–3209.

Monopolkommission (2018). Hauptgutachten XXII: Wettbewerb 2018. Hauptgutachten gemäß § 44 Abs. 1 Satz 1 GWB.

Normann, H.-T. and Sternberg, M. (2023). Human-algorithm interaction: Algorithmic pricing in hybrid laboratory markets, *European Economic Review* **152**: 104347.

Roth, A. E. and Murnighan, J. K. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma, *Journal of Mathematical Psychology* **17**(2): 189–198.

Rummery, G. A. and Niranjan, M. (1994). On-line q-learning using connectionist systems, *Technical report*, University of Cambridge, Department of Engineering.

Schwalbe, U. (2018). Algorithms, machine learning, and collusion, *Journal of Competition Law & Economics* **14**(4): 568–607.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences, *Machine Learning* **3**(1): 9–44.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*, Vol. 1, MIT Press, Cambridge.

von Schenk, A., Klockmann, V. and Köbis, N. (2023). Social preferences toward humans and machines: A systematic experiment on the role of machine payoffs, *Perspectives on Psychological Science*: Online First.

A Technical Details on Algorithmic Players

A.1 Human Imitator

The human imitator algorithm is designed to mimic the play in the Human treatment. At the beginning of a supergame, a human player from Human treatment with the same discount factor δ is drawn at random and the algorithm samples actions of this player throughout the entire supergame. At each stage game, the history of actions in the supergame up to this point is used by the algorithm to sample only from actions in comparable situations.

Suppose the human imitator algorithm plays a supergame and now has to choose an action in round t with history $H_t = \{(a_0, \bar{a}_0), \ldots (a_{t-1}, \bar{a}_{t-1})\}$, where $(a_{t'}, \bar{a}_{t'})$ denote the actions of the algorithm and its opponent, respectively, in period t'.

1. If history H_t has occurred to the sampled human player during the experiment once or multiple times, the algorithm samples an action out of all the actions taken by the human player at history H_t. That is, if the individual cooperated, for example, in 3 out of 4 observations with history H_t, the human imitator algorithm cooperates with 75% probability.

2. If the full history H_t has *not* occurred to the sampled human player, but the memory-1 history $(a_{t-1}, \bar{a}_{t-1})$ has occurred, the algorithm samples an action out of all the actions taken by the human player upon action profile $(a_{t-1}, \bar{a}_{t-1})$.

3. If the action profile $(a_{t-1}, \bar{a}_{t-1})$ has never occurred to the sampled human player during the experiment, the algorithm samples an action out of all the actions taken by *any* player in the experiment upon action profile $(a_{t-1}, \bar{a}_{t-1})$.

Figure A.1 illustrates the strategies employed by human imitators for different levels of discounting δ. One can clearly see that defection is the action played most often – for all three levels of discounting, but particularly for $\delta = 0.40$ and $\delta = 0.55$. For the largest continuation probability $\delta = 0.85$, mutual cooperation occurs significantly more often than for the other two values of δ, translating into larger expected payoffs.

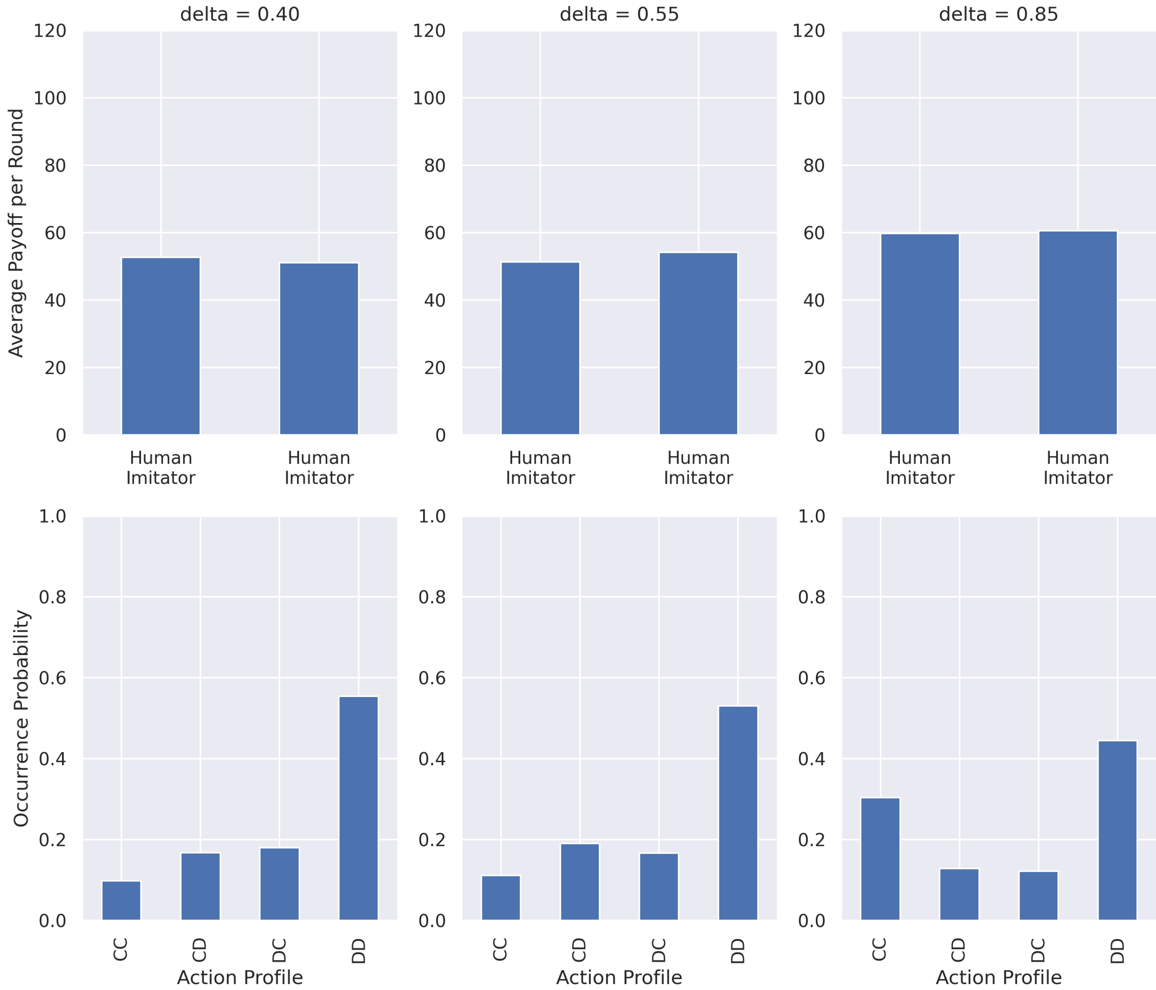

Figure A.1: Simulation Results Human Imitator vs. Human Imitator

Notes: The figure displays the empirical frequencies of action profiles when play between human imitators is simulated (1,000 times).

It is worthwhile to note that our implementation of the human imitator is model-free. Play from the human experiments is replicated statistically, by randomly drawing action choices of human subjects in comparable situations.

A.2 Optimized Reinforcement Learner

The optimized reinforcement learner algorithm is designed to maximize its own payoff when playing against a human imitator. We have striven for the most general implementation with minimal modeling assumptions. All relevant design choices have been made "objectively", by evaluating all choices and choosing the one that best accomplishes the goal of maximizing the algorithm's payoffs against human imitators.

When designing an algorithmic player, there are two types of modeling choices: first, choices concerning the richness of the strategies available to the algorithm and second, choices concerning the statistical optimization method that evaluates each strategy. Design choices concerning the richness of strategies (which corresponds to the specification of the "states" of the game that actions can be conditioned on) have a major impact on the play employed by the final algorithm and ultimately on its success in terms of payoff. As will be detailed in the following, we have chosen the state space such that the resulting algorithmic player performs best, without a priori assumptions. However, design choices concerning the statistical evaluation of strategies have no effect on the final algorithmic player – provided the statistical optimization method arrives at a global optimum of the payoff maximization. To ensure that our reinforcement learner arrives at a global optimum, we have run it with numerous starting values as well as random shocks. All meta-parameters are chosen so that the algorithm reliably arrives at the same optimum. Finally, in simulations against human imitators, we have verified that the final optimizer algorithm indeed earns maximal payoffs.

Our optimized reinforcement learner is based on a standard *Expected SARSA* algorithm. SARSA stands for State-Action-Reward-State-Action and describes the learning process of the algorithm: Starting in a given state, the algorithm tries an action, receives the immediate reward, finds itself in the next state, and evaluates the next action from there.

First, we need to define the state space. In case of a repeated Prisoner's Dilemma, the state space is a partition of the history of play. The optimized reinforcement learner can condition its action choice on the state, that is, on certain parts of the history of play.

Choosing a state space too small can restrict the algorithm to suboptimal action choices. Choosing the state space too large, on the other hand, can lead to identification problems when some states rarely occur and to convergence issues when unimportant parameters dilute the explanatory power of the model.

To select the optimal state space for the game at hand, the repeated prisoner's dilemma, we train the model with different state spaces up to size 65, let each trained model play against the human imitator and go with the model that achieves the highest payoffs. The optimal state space for our dataset turns out to be of size 17, including the action profiles of the previous period as well as the very first period of the supergame. This happens to be perfectly in line with the literature on the Repeated Prisoner's Dilemma, which has also found that the actions played in the previous period as well as the very first actions in the supergame provide the most information on subsequent actions (see, for example, Breitmoser (2015) and Fudenberg and Karreskog Rehbinder (2024)). Thus the state space consists of the following 17 states:

$$
\begin{aligned}
S \;=\; \big\{&0,\\
&CC\,CC,\; CC\,CD,\; CC\,DC,\; CC\,DD\\
&CD\,CC,\; CD\,CD,\; CD\,DC,\; CD\,DD\\
&DC\,CC,\; DC\,CD,\; DC\,DC,\; DC\,DD\\
&DD\,CC,\; DD\,CD,\; DD\,DC,\; DD\,DD\big\}
\end{aligned}
$$

where 0 denotes the very first period while the four letters stand for the own action in the initial period, the opponent's action in the initial period, the own action in the previous period and the opponent's action in the previous period, respectively.

Expected SARSA is a popular version of Q-learning. In Q-learning, every action in every state is assigned a Q-value. The Q-value $Q(s,a)$ determines the probability of playing the action a in the state s. During the learning process, Q-values are updated until convergence is achieved. The evolution of Q-values (i.e. learning) over periods t is updated using the following rule:

$$Q(s_t,a_t) \;\leftarrow\; (1-\alpha)\,Q(s_t,a_t) \;+\; \alpha\,\big(\pi(s_t,a_t)+\delta\,\mathbb{E}\big[Q(s_{t+1},a_{t+1})\big]\big) \tag{A.1}$$

where α is a meta parameter governing the learning rate, π denotes game payoffs and δ is the discount factor.

Mixed strategies $\sigma(s,a)$, i.e. the probability of playing action a in state s (from action set A_s), directly depend on Q-values and are given by the logit choice rule

$$\sigma(s,a) \;=\; \frac{e^{\gamma\,Q(s,a)}}{\sum\limits_{a'\in A_s} e^{\gamma\,Q(s,a')}} \tag{A.2}$$

where γ denotes the rationality parameter ($\gamma=0$ is random play, $\gamma\to\infty$ means always playing the action with the highest Q-value).

Training the algorithm requires that both the strategies and the Q-values converge. We start the learning process with uniform Q-values, random play ($\gamma = 0$) for initial experimentation, and high learning rate ($\alpha = 0.05$). From there, the algorithm plays and updates Q-values (and thus also strategies) repeatedly. After 100 supergames, γ is increased by 0.01 per supergame so that action choices become less random and stabilize. After 1,000 supergames, α is decreased at a rate of 0.005 per supergame so that the learning process converges. (Strategies must stabilize more quickly than Q-values.)

After 1,000 supergames, the algorithm starts checking for convergence. In each period, it compares the current strategies σ_t with the strategies σ_{t-100} 100 supergames ago. If the maximum difference in the strategy vectors is smaller than $1e-6$, the strategies are considered converged.

Figure A.2 depicts the evolution of strategies (probabilities of cooperation in the 17 states) as well as the Q-values of the 34 state-actions during the learning process of the reinforcement learning algorithm. After quite some initial experimentation, strategies and Q-values converge rather quickly – which is not surprising for this small game. In most states, optimal probabilities of cooperation are zero. In some states, probabilities remain between zero and one, indicating that actions the corresponding states do not have a large influence on final payoffs. Remarkably, for $\delta = 0.85$, a couple of cooperation probabilities converge to one, meaning that cooperation is worthwhile in these cases.

Figure A.3 illustrates the strategies employed by the reinforcement learner algorithm when playing against a human imitator algorithm as well as the corresponding outcomes, for different levels of discounting δ. Clearly, the reinforcement learner achieves higher payoffs than the human imitator in Figure A.1. For $\delta = 0.40$ and $\delta = 0.55$, it mainly achieves this via defection, translating into low payoffs for the human imitator opponent. For $\delta = 0.85$, the reinforcement learner achieves higher payoffs than for the other two values of δ, but mainly via increased cooperation, translating into even larger payoffs for the human imitator in this case.

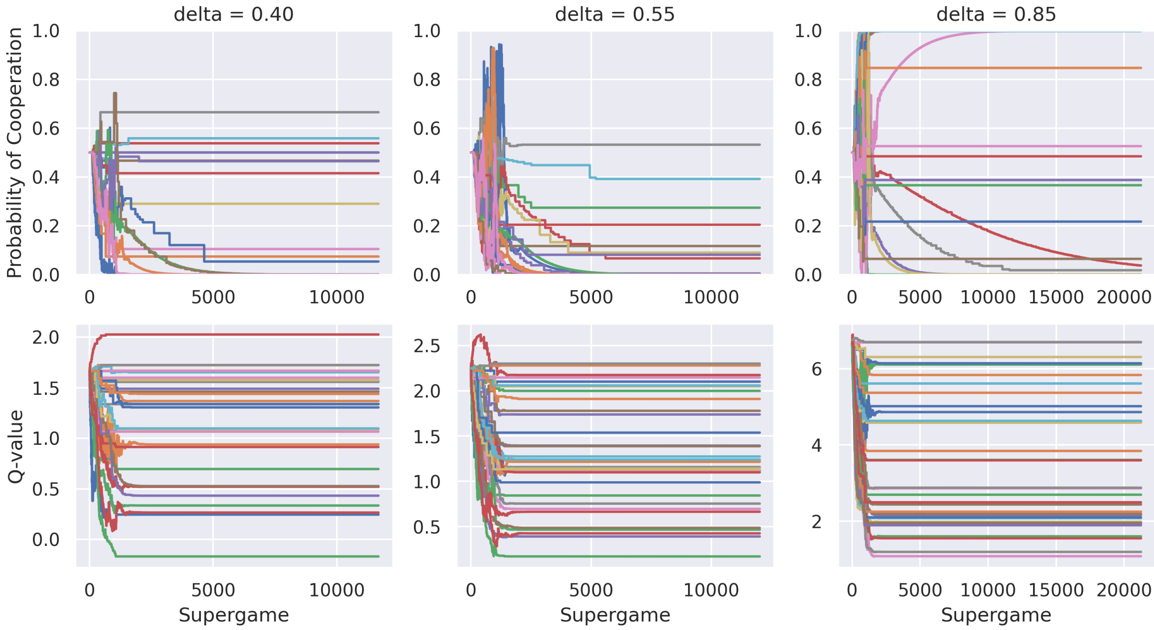

Figure A.2: Convergence of Optimized Reinforcement Learner

Notes: The figure displays the evolution of strategies as well as the corresponding Q-values during the learning process of the reinforcement learner algorithm.

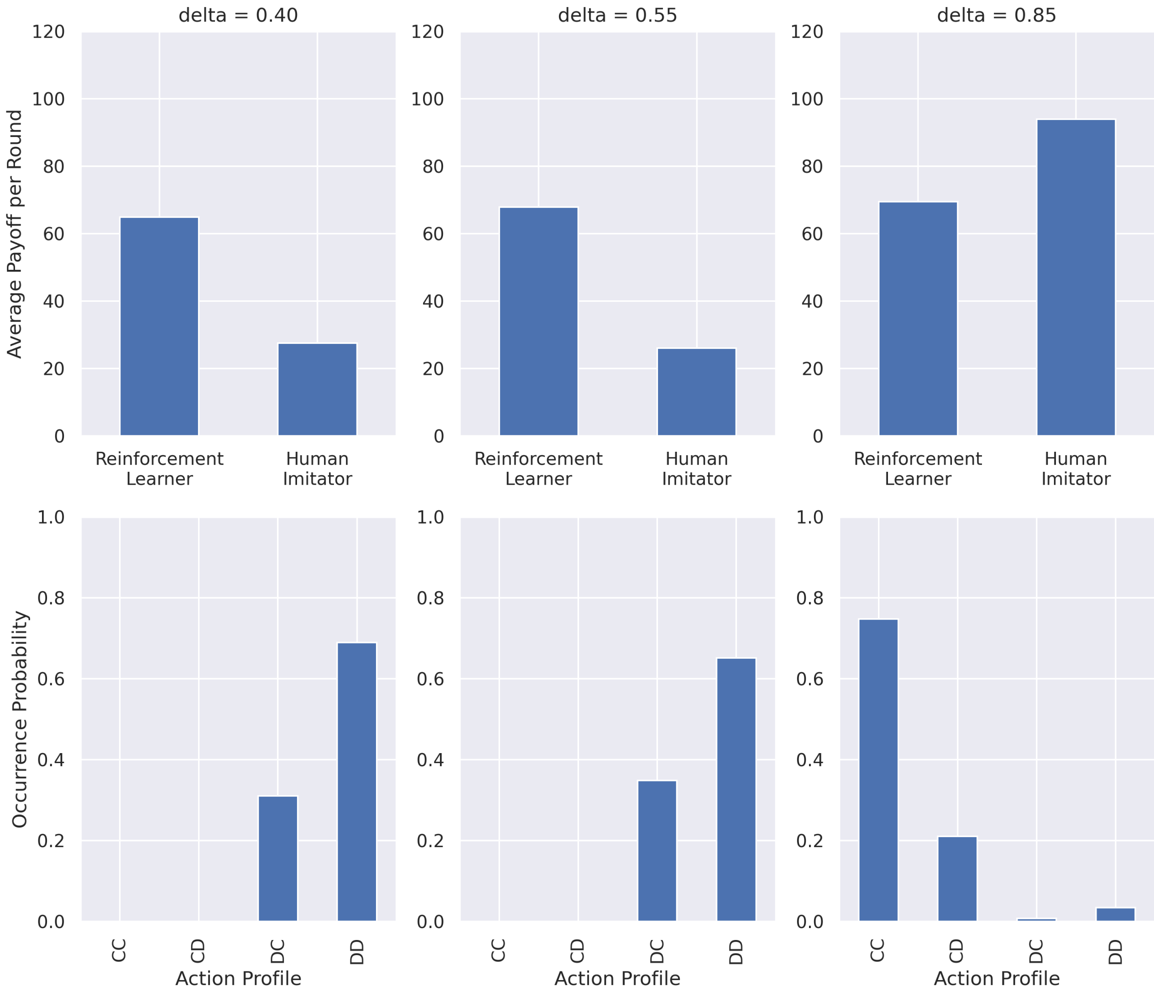

Figure A.3: Simulation Results Reinforcement Learner vs. Human Imitator

Notes: The figure displays the empirical frequencies of action profiles when play between reinforcement learners and human imitators is simulated (1,000 times).

B Additional Tables

Continuation Probability (δ)	**Opponent Type**		
	Human	Imitator	Optimal RL
0.40	316	156	154
0.55	312	150	160
0.85	294	150	155

Table B.1: Number of Participants

Notes: The table displays the number of participants, categorized by continuation probability and differentiated by the type and behavior of the opponent.

	(1)	(2)	(3)
δ	Human	Imitator	Optimizer
0.4	41.37%	26.19%	10.30%
0.55	53.19%	30.48%	8.04%
0.85	76.61%	65.33%	57.04%

(a) Initial Cooperation.

	(1)	(2)	(3)
δ	Human	Imitator	Optimizer
0.4	37.20%	23.98%	9.47%
0.55	45.91%	28.12%	6.50%
0.85	60.36%	50.21%	53.70%

(b) Overall Cooperation.

Table B.2: Human Cooperation Rates

Notes: The table reports average cooperation rates of human participants in the first period of each supergame (panel (a)) and across all rounds (panel (b)). The numbers always exclude the first three supergames.

	(1) Human	(2) Imitator	(3) Optimizer	(4) All
Supergame	-0.0182***	-0.0243***	-0.0339***	-0.0182***
	(0.00179)	(0.00271)	(0.00205)	(0.00179)
Imitator				-0.117***
				(0.0246)
Optimizer				-0.201***
				(0.0220)
Imitator × Supergame				-0.00611*
				(0.00324)
Optimizer × Supergame				-0.0157***
				(0.00272)
Constant	0.697***	0.579***	0.496***	0.697***
	(0.0139)	(0.0203)	(0.0171)	(0.0139)
N	9152	4560	4685	18397
R^2	0.011	0.020	0.044	0.078

Table B.3: Evolution of Initial Cooperation – Regressions

Notes: Regression results of initial cooperation on the supergame, treatment dummies, and interaction terms. Standard errors in parentheses, clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

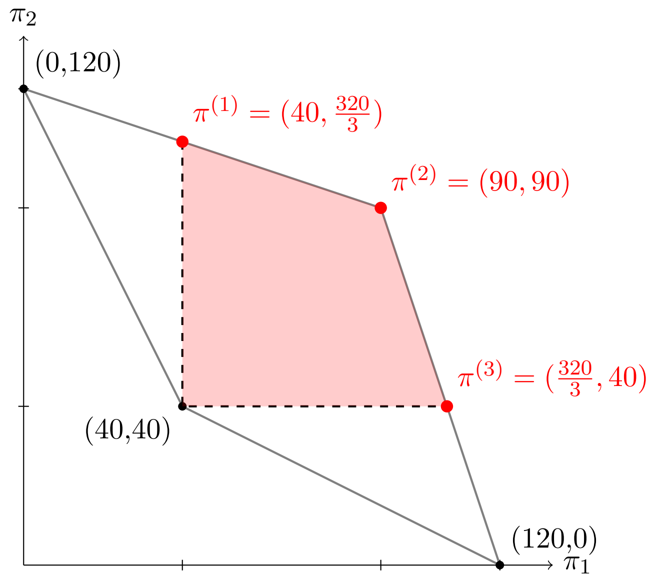

Figure C.1: Payoff Space

Notes: The figure shows the payoff space of the game. The marked area is the equilibrium payoff space and the marked points represent the vertices of the Pareto frontier.

C Robustness Check: Equilibrium Filtering

We follow the procedure by Blonski et al. (2011) to categorize the observed outcome paths in the repeated games into either equilibrium or non-equilibrium paths by employing their two-step filtering. First, we remove all paths that are not individually rational and violate

$$\sum_{t=1}^{T} \delta^{t-1} u_i(x_{1t}, x_{2t}) + \delta^T \pi_i \geq \frac{d}{1-\delta} \tag{C.1}$$

for $i = 1, 2$. Hereby, T denotes the length of the supergame, and the continuation payoffs π_i are elements of the equilibrium payoff space. In particular, we check whether condition (C.1) holds for at least one of the payoffs $(\pi_1, \pi_2) \in \{\pi^{(1)}, \pi^{(2)}, \pi^{(3)}\}$ forming the Pareto frontier of the equilibrium payoff space depicted in Figure C.1.

Second, we check a set of non-deviation conditions comparing the observed path with continued cooperation and deviations at any point in time $t \leq T$ with continued defection:

$$\begin{aligned} &\sum_{\tau=1}^{T} \delta^{\tau-1} u_i(x_{1\tau}, x_{2\tau}) + \delta^T \frac{c}{1-\delta} \geq \\ &\qquad \sum_{\tau=1}^{t-1} \delta^{\tau-1} u_i(x_{1\tau}, x_{2\tau}) + \delta^{t-1} u_i(\tilde{x}_{i,t}, x_{-i,t}) + \delta^t \frac{d}{1-\delta} \end{aligned} \tag{C.2}$$

	δ	(1) All	(2) Filter 1	(3) Filter 2
Human	0.4	41.37%	40.25%	36.12%
	0.55	53.19%	54.22%	53.11%
	0.85	76.61%	77.67%	77.89%
Imitator	0.4	26.19%	26.14%	22.94%
	0.55	30.48%	32.77%	31.31%
	0.85	65.33%	66.34%	66.35%
Optimizer	0.4	10.30%	6.52%	0.00%
	0.55	8.04%	5.12%	3.91%
	0.85	57.04%	63.64%	65.52%

Table C.1: Initial Cooperation Rates Across Opponents, Continuation Probabilities, and with Equilibrium Filtering

Notes: Column (1) includes all outcome paths, column (2) only those remaining after applying filter 1 (see (C.1)), column (3) only those remaining after applying filter 2 (see (C.2)).

for $i = 1, 2$ and $t \in \{1, \ldots, T\}$. Hereby, $\tilde{x}_{i,t}$ denotes the deviating action of player i in period t (i.e., C instead of D or D instead of C).

From the in total 13,821 outcome paths we collected across all opponents and all levels of δ (always excluding the first three supergames), 12,351 or 89.36% remain after filter 1 (89.38% for Human, 86.73% for Imitator, 91.91% for Optimizer) and 10,428 or 75.45% after filter 2 (74.17% for Human, 67.81% for Imitator, 84.14% for Optimizer).

Table C.1 reports initial cooperation rates across opponents and levels of δ for the whole sample, for paths remaining after filter 1, and for paths remaining after filter 2. For the highest continuation probability $\delta = 0.85 > \delta^*$, there are negligible changes in the estimated initial cooperation rate. For the two levels of δ below the threshold δ^* for which the δ^*-criterion predicts no cooperation, there is a decrease in the average likelihood of cooperating in the first round of the supergame for all opponent types. In particular, in the Optimizer treatment we observe (hardly) any initial cooperation after filter 2 (none for $\delta = 0.4$ and 3.91% for $\delta = 0.55$). As reported in panel (a) of Table C.2, the first dimension of δ^*-predictability

	(1) All	(2) Filter 1	(3) Filter 2
Human	0.118	0.140	0.170
Imitator	0.0429	0.0662	0.0837
Optimizer	−0.0226	−0.0140	0.0391

(a) Comparisons $\delta = 0.40$ vs. $\delta = 0.55$.

	(1) All	(2) Filter 1	(3) Filter 2
Human	0.294	0.305	0.324
Imitator	0.370	0.369	0.385
Optimizer	0.479	0.578	0.635

(b) Comparisons $\delta < \delta^*$ vs. $\delta > \delta^*$.

Table C.2: δ^*-Predictability Across Opponents and with Equilibrium Filtering

Notes: In both panels, column (1) includes all outcome paths (cf. Table 1), column (2) only those remaining after applying filter 1 (see (C.1)), column (3) only those remaining after applying filter 2 (see (C.2)).

hardly changes when applying the filtering. The second dimension remains mostly unaltered for human and human imitator opponents, while it strongly and significantly increases in the Optimizer treatment (all $p < 0.001$). Thus, through equilibrium filtering, the behavior of the human participants becomes even closer to the predicted strategies by the δ^*-criterion with defection for $\delta < \delta^*$ and cooperation for $\delta > \delta^*$.

Instructions

Introduction

In this study, you will be randomly and anonymously paired with another player.
Each of you simultaneously and privately decides whether you want to choose Action A or Action B.
Your payoffs will be determined by the choices of both, as illustrated below.

In each cell, the amount to the left is the payoff for you and to the right for the other player.

		The other player	
		Action A	Action B
You	Action A	90 points, 90 points	120 points, 0 points
	Action B	120 points, 0 points	40 points, 40 points

Rounds and Payoffs

You will play the described game repeatedly over several rounds.
The number of rounds you play is random.
Before every round, the computer determines randomly whether the game continues or ends.
You will learn the probability with which the game continues (**continuation probability**) before the first round.
It will remain the same for your whole experiment.
The higher the continuation probability, the higher the expected number of rounds you will play.

Your payoffs equal the sum of points you earn across all rounds.
Hereby, **100 points correspond to £0.10**.
On top of your earnings from the rounds, you receive a base pay of £2.50.

Try it out!

Select **one action for you** and **one for the other player**, i.e. for both players.
You will then learn what would happen in such a scenario.

		The other player	
		Action A	Action B
You	Action A	90 points, 90 points	0 points, 120 points
	Action B	120 points, 0 points	40 points, 40 points

Suppose that you chose **Action A**.
Suppose that the other player chose **Action A**.
In this case, you would get **90 points**, the other player would get **90 points**.

Comprehension Questions *[correct answers marked with filled square]*

Please answer the following comprehension questions.
You need to answer all of them correctly.
You have **3 out of 3 attempts left** to answer all correctly.

Do you play the game for a *fixed or random* number of rounds?
- ☐ The number of rounds is fixed.
- ■ The number of rounds is random.

Do you play the same game *repeatedly* or does the game *change between rounds*?
- ■ The game remains the same.
- ☐ The game changes between rounds.

If the continuation probability increases, how does the *expected number of rounds* you play change?
- ■ The expected number of rounds increases.
- ☐ The expected number of rounds decreases.
- ☐ There is no effect on the expected number of rounds.

Is the probability that after the first round the second starts *smaller / equal / larger* than the probability that after the fifth round the sixth starts?
- ☐ Smaller
- ■ Equal
- ☐ Larger

Is the expected number of rounds that the game will keep going *smaller / equal / larger* after the first round than after the fifth round?
- ☐ Smaller
- ■ Equal
- ☐ Larger

Comprehension Questions

Please answer the following comprehension questions using the game table below.
You need to answer all of them correctly.
Fill out the entry fields and click on the = button in the respective line to calculate the result.
You have **3 out of 3 attempts left** to answer all correctly.

		The other player	
		Action A	Action B
You	Action A	90 points, 90 points	0 points, 120 points
	Action B	120 points, 0 points	40 points, 40 points

Your action:	**Action A**
The other players' action:	**Action A**
Your payoff:	**90 points**
The other players' action:	**90 points**

Scenario 1

Consider the following scenario.
The game lasts **7 more rounds**. *[δ=0.4, 0.55: 2 rounds; δ=0.85: 7 rounds]*
In the current round, you and the other player both choose **Action A**.
In the future 6 rounds, you and the other player continue to choose **Action A**.
What is your total payoff?

Payoff in current round		Payoff in future rounds		Remaining future rounds		Total payoff
90 points	+	90 points	×	6	=	630 points

Scenario 2

Consider the following scenario.
The game lasts **7 more rounds**. *[δ=0.4, 0.55: 2 rounds; δ=0.85: 7 rounds]*
In the current round, you choose **Action A** and the other player chooses **Action B**.
In the future 6 rounds, you and the other player both always choose **Action B**.
What is your total payoff?

Payoff in current round		Payoff in future rounds		Remaining future rounds		Total payoff
0 points	+	40 points	×	6	=	240 points

Scenario 3

Consider the following scenario.
The game lasts **7 more rounds**. *[δ=0.4, 0.55: 2 rounds; δ=0.85: 7 rounds]*
In the current round, you choose **Action B** and the other player chooses **Action A**.
In the future 6 rounds, you and the other player both always choose **Action B**.
What is your total payoff?

Payoff in current round		Payoff in future rounds		Remaining future rounds		Total payoff
120 points	+	40 points	×	6	=	360 points

Assessment 1

The main part of the experiment is now over.
Before you finish, we ask you a few questions.

Which of the following goals describes best your chosen actions in the games?

O Maximizing own payoff
O Maximizing cooperation
O Achieve equal payoffs
O Maximizing the sum of payoffs
O Minimizing your opponent's payoff

By answering the following question, you can earn an additional bonus.
We compare your answer the answer of your opponents in the previous games.
If your answer matches the majority of your opponents' answers, you will earn a bonus of 200 points.

Which of the following goals do you think describes best the majority of your opponents' chosen actions in the game?

O Maximizing own payoff
O Maximizing cooperation
O Achieve equal payoffs
O Maximizing the sum of payoffs
O Minimizing your payoff

Assessment 2

Before showing you the final results, we ask you another question.
By answering this questions, you can earn an additional bonus.
If your answer is correct, you will earn a bonus of 200 points.

Against what type of opponent do you think you have played?

O Another human participant
O An intelligent algorithmic player
O A randomly playing bot

Survey

Please answer the following questions.

What is your age?

What is your gender?

O Male
O Female
O Other

What is your highest educational degree?

O No degree
O High School
O Bachelor
O Master
O PhD

If you go/went to university, what is/was your major?

- O Not applicable
- O Economics
- O Law
- O Psychology
- O Political sciences
- O Medicine
- O Natural sciences
- O Mathematics
- O Engineering
- O Other social sciences
- O Other

How familiar are you with artificial intelligence and machine learning?

- O Not familiar at all
- O Rather not familiar
- O Neutral
- O A little familiar
- O Very familiar

How much confidence do you have in new technologies like artificial intelligence?

- O No confidence at all
- O Rather no confidence
- O Neutral
- O A little confidence
- O Strong confidence

Please describe your strategy in the experiment in about two sentences.