How Intelligent Players Teach Cooperation

Eugenio Proto *

Aldo Rustichini†

Andis Sofianos‡§

September 5, 2019

Abstract

We study cooperation rates in an infinitely repeated Prisoner's Dilemma game focusing on whether and how subjects of higher intelligence affect the behavior of others. Proto et al. (2019) establish that participants of higher intelligence have a higher cooperation rate than otherwise similar participants of lower intelligence if playing in two separate groups (*split* treatment). Here we study how cooperation rates change over time in mixed groups (*combined* treatment).

The first main finding is that the cooperation rate in the combined treatment is substantially higher than the rate in lower intelligence groups (with IQ 76-106), and slightly lower than in higher intelligence groups (with IQ 102-127) in the split treatment. Teaching subjects could be more forgiving with the aim to help others understand that it is in their best interest to mutually cooperate (which would be a form of *intentional* teaching). On the other hand, teachers could just consistently best respond to their beliefs – for example by punishing any deviation – hence, providing an example of behaviour for the others to learn to play more efficiently (a form of *unintentional* teaching).

We find support for the latter hypothesis: higher intelligence players use retaliatory strategies more often when they are in the combined treatment. They do so more consistently than lower intelligence players, enforcing punishment when it is due, thus providing the low intelligence players in later encounters with the appropriate incentives to cooperate.

JEL classification: C73, C91, C92, B83
Keywords: Repeated Prisoner's Dilemma, Cooperation, Intelligence, Teaching Cooperation

*University of Bristol, CEPR, IZA and CesIfo; email: eugenio.proto@gmail.com
†University of Minnesota; email: aldo.rustichini@gmail.com
‡University of Heidelberg; email: andis.sofianos@uni-heidelberg.de
§The authors thank several colleagues for discussions on this and related research, in particular Pierpaolo Battigalli, Maria Bigoni, Marco Casari, Guillaume Frechette, David Gill, Marco Lambrecht, Fabio Maccheroni, Salvatore Nunnari, Joerg Oechssler, Emel Ozbay, Erkut Ozbay, Andy Schotter. The conference and seminar participants at the University of Bologna, University of Siena, Bocconi University, Purdue University, University of Maryland and the HeiKaMaX. The University of Heidelberg provided funding for this research. AR thanks the National Science Foundation, grant NSF 1728056.

1 Introduction

An important characteristic determining the levels of cooperation in different strategic environments is the intelligence of players. Recent literature has begun to investigate the way in which different cognitive skills affect learning in strategic environments (e.g. Gill and Prowse, 2016; Alaoui and Penta, 2015) while some recent findings indicate that lower intelligence is associated with more reliance on social learning through imitation (e.g. Muthukrishna et al., 2016; Vostroknutov et al., 2018). Proto et al. (2019) find that in a *split treatment* where the subjects are allocated into two groups on the base of their intelligence, only the higher intelligence groups converge to full cooperation. This result identifies an important factor affecting cooperation, however, such separation of individuals in distinct classes does not occur in everyday life and the question whether and how a group influences the other is left open.

To tackle this we adopt an experimental design where such intelligence separation does not occur. We find strong evidence that less intelligent people learn how to profitably cooperate from more intelligent.[1] Specifically, we observe an important change in the frequency of strategies in the combined treatment as compared to the treatment with high Intelligence players only. We see a shift of the frequency strategies played in the direction of less lenient strategies.

We argue that this is an instance of a general phenomenon: if the fraction of players with limited cognitive skills, and thus the average probability of errors increases, they will be, broadly speaking, led to adopt stricter strategies. In order to understand this complex mechanism of learning and teaching, we analyze a model, where differences among players or among groups as differences in the working memory. A lower working memory produces a larger probability of error in implementing the strategy. We focus here not on the errors in the choice of action, but on errors in the management of the strategy. We model the strategy as an automaton, and the essential part of the management of the strategy is to correctly choose the next state in the automaton, given the current state and the observed action profile. We assume that a lower working memory ability produces more frequent errors in this management. We study the effect on the frequency of strategies in the population at the evolutionary equilibria of two different benchmark models (the proportional imitation model (Schlag (1998)) and the best response (Gilboa and Matsui (1991))).

Establishing cooperation under repeated interactions is a complex process as many factors are involved and many skills are necessary. Even in a stylized experimental setting, players need to choose the right strategy, have correct beliefs about the opponent and be able to follow coherently a strategy after it has been chosen. Yet, experimental evidence (e.g. Dal Bó, 2005; Dreber et al., 2008; Duffy and Ochs, 2009; Dal Bó and Fréchette, 2011; Blonski et al., 2011) show how subjects, when gains from cooperation are sufficiently large, tend to cooperate under repeated interactions. Fudenberg et al. (2012) analyse the effect of uncertainty in the implementation of different strategies in games of cooperation under repeated interactions, and show how subjects factor in this noise when playing and become more lenient and forgiving. In our setting the noise is endogenous, and is mostly due to the mistakes of the less intelligent. We show that the more intelligent being less lenient and forgiving than the less intelligent reduces the noise and improves efficiency and payoffs.

There is evidence that subjects teach other subjects how to play efficiently. In particular Hyndman et al. (2012) show that some participants act as teachers and play a forward looking strategy trying to influence the action of other players. Their design adopts a finitely repeated game, where the stage game has a unique pure strategy Nash equilibrium with an outcome on the

[1] Cooperation rates in the combined treatment increases for the lower intelligence players, and slightly decreases for higher intelligence ones. Specifically, the cooperation rate is substantially higher for lower intelligence players (those with IQ in the range 76-106) when we compare them to the split treatment. Instead, the cooperation rate is slightly lower for the higher intelligence players (range IQ 102-127), again compared to the split treatment.

Pareto frontier. They show that subjects do not best respond to their beliefs about the choice of the other players, presumably with the intent of teaching the others (*active* teaching).

However, in the repeated Prisoners' Dilemma, subjects have a short-term incentive to deviate from the cooperative outcome, thus other mechanisms of teaching are possible. Specifically, Proto et al. (2019) show that the existence of a tension between short-term and long-term objectives leads the less intelligent to inefficiently reach non-cooperative outcomes. Along the lines of the active teaching hypothesis, teaching subjects in our setup could become more forgiving, and hope that other subjects understand that is in their best interest to mutually cooperate. What we find, instead, is that they just consistently best respond to their beliefs – for example by punishing any deviation – providing an example of behaviour for the others to learn to play more efficiently; we call this *passive* teaching.

The paper is organized as follows: In section 2 we formulate the broad question on the effect of mixing subjects with different levels of intelligence, and present the experimental design. We then show (subsection 2.2) how cooperation rates are affected by he intelligence composition of the groups. In section 3 we present the main result of the model investigating the frequency of strict and lenient strategies in populations with different error rates. In section 4 we investigate the difference in the process of belief updating and learning the best response across the different groups and treatments. Section 5 presents our conclusions. Additional technical analysis, robustness checks, details of the experimental design and descriptive statistics are in the appendix.

2 The experiment and initial hypothesis

Our design involves a two-part experiment administered over two different days separated by one day in between. Participants are allocated into two groups according to cognitive ability that is measured during the first part, and they are asked to return to a specific session to play several repetitions of a repeated game. Each repeated game is played with a new partner. We have two treatments: one where participants are separated according to cognitive ability and one where participants are allocated into sessions where cognitive ability is similar across sessions. We call the former the *IQ-split* treatment and the latter the *Combined* treatment. The subjects were not informed about the basis upon which the split was made.[2]

In an environment where cooperation can be sustained as a subgame perfect equilibrium, our first question examines how cooperation rates and profit of the players compare while in combined groups or in separate groups. We will see in particular that less intelligent players learn to cooperate and to play more efficiently in the combined treatments and the average payoffs are higher in the combined treatments than in the split treatments taken together. We will also see that less intelligent subjects commit more mistakes than the more intelligent ones when they play separately, but not when they play together with the more intelligent.

On the basis of these observations, in section D we will analyse a model of learning showing why the less intelligent learn when they play with the more intelligent, then in the next section, we will test the prediction of the model using the data of the experiment.

[2]During the de-briefing stage we asked the participants if they understood the basis upon which the allocation to sessions was made. Only one participant mentioned intelligence as the possible determining characteristic.

2.1 Experimental Details

Day One

On the first day of the experiment, the participants were asked to complete a Raven Advanced Progressive Matrices (APM) test of 36 matrices. They had a maximum of 30 minutes for all 36 matrices. Before initiating the test, the subjects were shown an example of a matrix with the correct answer provided below for 30 seconds. For each item a 3×3 matrix of images was displayed on the subjects' screen; the image in the bottom right corner was missing. The subjects were then asked to complete the pattern choosing one out of 8 possible choices presented on the screen. The 36 matrices were presented in order of progressive difficulty as they are sequenced in Set II of the APM. Participants were allowed to switch back and forth through the 36 matrices during the 30 minutes and change their answers.

The Raven test is a non-verbal test commonly used to measure reasoning ability and general intelligence. Matrices from Set II of the APM are appropriate for adults and adolescents of higher average intelligence. The test is able to elicit stable and sizeable differences in performances among this pool of individuals. This test was among others implemented in Proto et al. (2019) and Gill and Prowse (2016) and has been found to be relevant in determining behaviour in cooperative or coordinating games.

Subjects are usually not rewarded for completing the Raven test. It has though been reported that Raven scores slightly increase after a monetary reward is offered to higher than average intelligence subjects (e.g. Larson et al., 1994). With the aim of measuring intelligence with minimum confounding with motivation, we decided to reward our subjects with 1 Euro per correct answer from a random choice of three out of the total of 36 matrices. During the session we never mentioned that Raven is a test of intelligence or cognitive abilities.

Following the Raven test, the participants completed an incentivised Holt-Laury task (Holt and Laury, 2002) to measure risk attitudes. Finally, participants were asked to respond to a standard Big Five personality questionnaire together with some demographic questions, a subjective well-being question and a question on previous experience with a Raven's test. No monetary payment was offered for this section of the session and the subjects were informed about this. We used the Big Five Inventory (BFI); the inventory is based on 44 questions with answers coded on a Likert scale. The version we used was developed by John et al. (1991) and has been recently investigated by John et al. (2008).

All the instructions given on the first day are included in the supplementary material.[3]

Day Two

On the second day, the participants were asked to come back to the lab and they were allocated to two separate experimental sessions. The basis of allocation depends on the treatment. In the IQ-split treatment, participants were invited back according to their Raven scores: subjects with a score higher than the median were gathered in one session, and the remaining subjects in the other. We will refer to the two sessions as *high-IQ* and *low-IQ* sessions.[4,5] In the combined treatment, we made sure to create groups of similar Raven scores across sessions. To allocate participants

[3] This is available online at X

[4] The attrition rate was small, and is documented in table A.1.

[5] In cases where there were participants with equal scores at the cutoff, two tie rules were used based on whether they reported previous experience of the Raven task and high school grades. Participants who had done the task before (and were tied with others who had not) were allocated to the low-IQ session, while if there were still ties, participants with higher high school grades were put in the high session.

to second day sessions, we ranked them by their Raven scores and split by median. Instead of having high- and low-IQ groups though, we alternated in allocating participants in one session or the other.[6]

The task they were asked to perform was to play an induced infinitely repeated Prisoner's Dilemma (PD) game. Table 1 reports the stage game that was implemented.

Table 1: **Prisoner's Dilemma.** C: Cooperate, D: Defect.

	C	D
C	48,48	12,50
D	50,12	25,25

We induced infinite repetition of the stage game using a random continuation rule: after each round the computer decided whether to finish the repeated game or to have an additional round depending on the realization of a random number. The continuation probability used was $\delta = 0.75$. We used a pre-drawn realisation of the random numbers; this ensures that all sessions across both treatments are faced with the same experience in terms of length of play at each decision point. As usual, we define as a supergame each repeated game played; period refers to the round within a specific supergame; and, finally, round refers to an overall count of number of times the stage game has been played across supergames during the session. The length of play of the repeated game during the second day was either 45 minutes or until the 151st round was played depending on which came first.

The parameters used are identical to the ones used by Dal Bó and Fréchette (2011) and Proto et al. (2019). The payoffs and continuation probability chosen entail an infinitely repeated Prisoner's Dilemma game where the cooperation equilibrium is both subgame perfect and risk dominant.[7]

The matching of partners is done within each session under an anonymous and random rematching protocol. Partipicants played as partners for as long as the random continuation rule determines that the particular partnership is to continue. Once each match was terminated, the subjects were again randomly and anonymously matched and started playing the game again according to the respective continuation probability. Each decision round for the game was terminated when every participant had made their decision. After all participants made their decisions, a screen appeared that reminded them of their own decision, indicated their partner's decision while also indicated the units they earned for that particular round. The group size of different sessions varies depending on the numbers recruited in each week.[8] The participants were paid the full sum of points they earned through all rounds of the game. Payoffs reported in table 1 are in terms of experimental units; each experimental unit corresponded to 0.003 Euros.

Upon completing the PD game, the participants were asked to respond to a short questionnaire about any knowledge they had of the PD game, some questions about their attitudes towards cooperative behaviour and some strategy-eliciting questions.

Implementation

The recruitment was conducted through the Alfred-Weber-Institute (AWI) Experimental Lab subject pool based on the Hroot recruitment software (Bock et al., 2014). All sessions were administered at the AWI Experimental Lab in the Economics Department of the University of Heidelberg. A

[6]Again, the attrition rate was small, and is documented in table A.2.

[7]See Dal Bó and Fréchette (2011), p. 415 for more details

[8]The bottom panels of tables A.1 and A.2 in the appendix list the sample size of each session across both treatments.

total of 214 subjects participated in the experimental sessions. They earned on average around 23 Euros each; the show-up fee was 4 Euros. The software used for the entire experiment was *Z-Tree* (Fischbacher, 2007).

We conducted a total of 8 sessions for the IQ-split treatment; four-high IQ and four low-IQ sessions. There were a total of 108 participants, with 54 in the high-IQ and 54 in the low-IQ sessions. For the combined treatment we conducted a total of 8 sessions with a total of 106 participants. The dates of the sessions and the number of participants per session, are reported in tables A.1 and A.2 in the appendix. The recruitment letter circulated is in the supplementary material.[9]

2.2 Cooperation rates, payoffs

We start by comparing cooperation rates and payoffs across the two treatments for the two intelligence groups. Figure 1 shows that subjects increasingly choose cooperation as first move across all treatments. Subjects in the High IQ sessions converge faster to almost a full cooperation rate, while in the low IQ sessions this pattern is slower; subjects in this group converge to a cooperation rate smaller than 100 per cent (left panel). Subjects in the combined session do better than in the low sessions and worse than the ones in the high IQ session (right panel).[10] These results replicate the findings in Proto et al. (2019) by using a different subject pool in a different country.

Moreover, table 2 shows that in the high IQ sessions subjects earn 2.5 units and cooperate 10% more than in the combined sessions in the first 20 supergames, while in low IQ sessions they cooperate about 20% and earn 5.5 units less than in the combined sessions. After the 20th supergame, there is no longer a significant difference between high IQ and combined sessions. This suggests that low IQ learn to play as efficiently as the high IQ in the second part of the sessions in the combined treatment. Meanwhile in the low IQ sessions the differences in both cooperation and payoffs remain constant. Table 3 shows that IQ is not significant in determining cooperation in the first round in either of the two treatments, which suggests that the differences in cooperation between individuals with different cognitive skills are only due to the learning effect during the sessions.[11] Interestingly, risk aversion is the only significant determinant of cooperation at the beginning of each session; hence learning in different environments seem key in determining the level of cooperation.

Figure 2 shows that the average payoff per interaction is consistently higher in the combined sessions than in the low IQ sessions, showing that in the combined treatment subjects play in average more efficiently than in the split treatments.[12]

This preliminary evidence shows that subjects in general learn to cooperate on the basis of their cognitive skills, but also of the distribution of the cognitive skills within the group, so a complex mechanism of learning is in place. The following model shows this in detail.

3 A Model of Errors and Strategy Evolution

In the following we model differences among players or among groups as differences in the working memory. A lower working memory produces a larger probability of error in implementing the strategy. We focus here not on the errors in the choice of action, but on errors in the management

[9] See note 3.

[10] In figure A.4 of the appendix we present the cooperation rates by session.

[11] Proto et al. (2019) finds a similar pattern in the laboratory at the University of Warwick

[12] This is also seen in table A.3 in the appendix where total earnings as well as average payoff per round are significantly higher in the combined sessions than in the split sessions.

of the strategy. We model the strategy as an automaton, and the essential part of the management of the strategy is to correctly choose the next state in the automaton, given the current state and the observed action profile. We assume that a lower working memory ability produces more frequent errors in this management. We study the effect on the frequency of strategies in the population at the evolutionary equilibria of two different benchmark models (the proportional imitation model (Schlag (1998)) and the best response (Gilboa and Matsui (1991))). We call A the *always defect* strategy, G the *grim trigger*, and T the *Tit-for-Tat*.

The main results of the analysis (developed in detail in section D of the Appendix) are the following:

1. When players have perfect working memory, equilibria of the strategy choice game, and steady states of the learning process and the evolutionary model is indeterminate;

2. When errors occur, even of arbitrarily small side, then there are three locally unique, and locally stable steady states corresponding to the pure strategies of the game in which players choose strategies in the repeated game. There are overall seven steady states, three stable, one unstable, and three saddle points;

3. Thanks to the previous result, when errors are positive (however small) we can define basins of attraction for each of the strategies, thus providing a theoretical basis to predict relative frequency of strategies as function of the error rate;

4. As ϵ becomes larger (that is, as more error prone the group of players is), the basin of attraction of the stricter strategies become larger: the size of the basin of the A strategy becomes larger than that of G and T combined, and that of G becomes larger than that of T;

5. Which strategies survive depends on δ; in all cases, for low level or errors all strategies may survive depending on the initial condition; and for low level of error only defection survives. For intermediate levels, the two surviving strategies are defect and grim trigger for low δ, and defection and Tit-for-Tat for high δ (see conclusion I.5);

6. As δ tends to 1, for fixed error, the opposite happens: the basin of attraction of the G and T strategies becomes larger; and when strategies are limited to $\{G, T\}$, that of T increases to cover the entire interval.

7. The results described so far are independent of the specific model of evolution we adopt. In the following we compare the evolutionary dynamics with two different models, Proportional imitation or best response, and find that they are qualitatively similar.

8. In section (D) of the appendix we develop a model of learning in a population of players with heterogeneous beliefs, who hold and update beliefs as in the model underlying our data analysis, and show that the resulting dynamics is close to that described by one of the evolutionary models (the best response dynamics).

3.1 A Simple Illustration

A first intuitive understanding of the way in which the error rate affects the basins of attraction of the three strategies can be obtained by considering Figure (10). This figure reports the phase portrait of the vector field for the Best Response dynamics, at different error probability, ranging from 1 per cent (top left panel) to 25 per cent (bottom right panel). This is the range of errors that is most relevant for our purposes, since it corresponds to the range or error we observe in our data. Payoff in the stage game are set as in our experimental design (see section (E) below for details). The discount factor (or equivalently, the continuation probability) is the same as in our experimental design, namely 75 per cent.

The triangle in each panel of the figure is the two-dimensional projection of the simplex, and each point in the triangle represents points (pG, pT) such that $pG + pT \leq 1$. Thus, the frequency of the strategy (A, G, T) is equal to $(1 - pG - pT, pG, pT)$. The lines in the triangular regions represent the isoclines, namely the set of points at which the time derivatives of the two variables is equal to zero. [13] The red line indicates the set of points where $\frac{dpT}{dt} = 0$, the blue $\frac{dpG}{dt} = 0$. These lines split the triangular region into three subsets, each one containing the point corresponding to a pure strategy. In each of these regions the fraction of the attracting strategy increases over time, the fraction of the other to decreases. The interior of each of these three subsets consists entirely of points that are attracted to the pure strategy that is contained in the boundary of the subset. For example, the basin of attraction of A is the smaller region (also triangular) at the bottom left.

The fact that the boundaries of the regions are straight-lines follows from the special nature of the Best Response dynamics. For comparison, the reader can compare this figure with figure (A.10), reporting the same results for the Replicator Dynamic (or proportional Imitation).

Several points that appear clearly in Figure (10) are worth pointing out. First, for all error rates, each of the three strategies has a basin of attraction in the interior of the triangular region: in other words, all strategies survive in the range of error rate we observe in the data. Second, the size of the region attracted to A increases monotonically with the error rate. Finally, if we consider the complementary region of points that are *not* attracted to A, but to G or T, we note that the relative fraction that goes to G (the second strict strategy) is increasing as the error rate is increasing. Note that as a result, whereas at the lowest error rate (top left) the T region is the largest, at the highest (bottom left) it is the smallest. In few simple words, *as the error rate increases strategies become more frequent precisely in strictness order, with stricter strategies becoming more frequent.*

We now proceed to present how the analysis of learning can be implemented.

Accordingly, table 4 shows that– in the first 20 supergames– subjects in the high IQ sessions and in the combined sessions increasingly open with cooperation at faster speed than in the combined treatment, while in the low IQ session the cooperation increase less than in the combined sessions (columns 1 and 2); subjects in the low IQ sessions tend to catch-up with the others in the in the second part (column 3 and 4).

This represents evidence that lower IQ learn to play more cooperative strategies when mixed with higher IQ faster than when they play together. This is confirmed by tables 5 and 6, where we estimated the probability of the different strategies subjects adopt in the split treatments.[14] In table 5 we note that high IQ play always defect (then open with a defection) the 20% of time in the first 5 supergames, this probability goes essentially to zero in the last 5 supergames. For the the low IQ, the probability of always defect is close to 50% in the first 5 supergames but is still sizable in the last 5 supergames (about 24%). This behaviour needs to be compared with the strategies played by the low IQ groups in the combined sessions– displayed in table 6. Already, in the 1st 5 supergames low IQ play non cooperative strategy about 19% of time and this probability decreases even more in the last 5 supergames (about 13%), so they learn quickly to cooperate when they play with the high IQ in the combined treatments.

[13] More precisely, since we later define the Best Response dynamics as a differential inclusion, the set of points at which zero is the set of derivatives; in the discretization used to produce figure (10), the distinction between differential inclusion and differential equation is irrelevant.

[14] We followed Dal Bó and Fréchette (2011) for this estimation, the details are provided in the online appendix of their paper.

Why this increase in cooperation in the opening rounds in session where there is a larger number of high IQ subjects occurs?

If this increase in cooperation is driven by subjects beliefs about the other subjects play in period 1 of each supergame, we should observe that subjects whose partners opened with cooperation in the past more often should should increase the probability of opening with cooperation in the present. This is what we observe in column 2 and 4 of table 4, where the coefficients of the partners' pasts cooperation in the 1st rounds are positive and significant. In column 2, however, the other coefficients show a positive trend and a significant difference in the sessions with more high IQ subjects, suggesting that the increase in 1st period cooperation is not entirely due to the past behaviour of the subjects in period 1.[15]

Figures 5, 6 and 7 provide further insights about this learning process. In figure 5, it is evident that high IQ learn to reciprocate with cooperation faster than low IQ, and low IQ players learn faster when they play in the mixed sessions and can observe that high IQ are more keen to reciprocate cooperation. This learning mechanism is even clearer in figures 6 and 7, where we report the percentage of defections a subject face in period t+1 after cooperating at t: this is a situation that clearly discourages cooperation. For the high IQ (figure 6) there is a declining pattern both when they play separately and when they play combined with low IQ. Although there seems to be a difference when they play separately in the first part of the sessions. In the second part the rate of defection after cooperation is similar whether high IQ play separately or not, suggesting a learning pattern for the low IQ. From figure 7, we observe that the low IQ face defection after cooperation more when they play separately then when they play combined and the decline of this pattern when they play separately is less clear than when they play combined with the high IQ. Therefore, it is also the behaviour of the partners after cooperation rather than the 1st period cooperation which seems to drive the patterns of cooperation in the different groups and treatments that we observe in figure 1.

But are the low IQ learning to play the optimal strategy when they are mixed with the high IQ? In other words, is opening with cooperation an optimal choice from the low IQ point of view? In tables 7 and 8 we analyze the optimality of the the most frequent strategies subjects plays in the different treatments. As we can see in table 7, when they play in the split sessions, the low IQ play Grim, Tit for Tat, Win stay Lose Shift (that we define "sophisticated cooperation" (SC)) only an estimated 29% of time, despite this being the strategy yielding the highest expected payoffs (38.19). From table 8 we notice that in the combined treatment the probability of SC is already about 36% in the 1st 5 supergames and goes up until 81% in the last 5 supergames. *Hence, the low IQ learn to play more optimal strategies when mixed with the high IQ*

3.2 The teaching of the high IQ subjects

Less intelligent learn to play more efficient when mixed with high intelligent, our last question is then: *How the two groups (i.e. more or less intelligent) change their strategies when the play combined and when they play separately?*.

As we said in reference to figure 6, high IQ face defection after cooperation more often when they play in the mixed sessions, but the likelihood of this situation decline in the second part of this session. How this decline has been achieved? if the high IQ is patient and forgives the 1st defection, will the partner revert back to cooperation?

[15] In section 4, we will formally investigate the process of beliefs formation and updating for the first periods of each supergame.

Figure 8 suggests that this is NOT the case. After being suckered, high IQ face more defection when they play in the combined session than when they play separately, especially in the second part of the session. High IQ, when matched with other subjects with similar IQ, after being the sucker in the last period their partners revert to cooperation with 50% chance. On the other hand, in a combined session this chance declines to less than 25%. Patience then seems to pay more when high IQ are matched with other high IQ, but less so when they are playing in combined groups (i.e. with possible low IQ partners also). An effect of this can be observed in table 5, the high IQ play AC about 35% of time in the last 5 supergames when the play separately. While, in the combined sessions, AC is essentially 0 as we can see in table 6.

In what comes next, we analyse this change of behaviour between combined and split treatments in a more systematic way. There is widespread evidence that subjects overwhelmingly play *memory one* strategies in repeated prisoners' dilemma games (see Dal Bó and Fréchette (2018)). This is consistent with the results presented in tables 5 and 6, which show that AC, AD, Grim Trigger and TfT cover between 85% and 100 % of the strategies played by all subjects. As in the previous simulation we assume that every subject choose a strategy that applies for a number of supergames. Accordingly, we assume that choices in every round are determined by the past outcome (hence by both his and his partner's choice in the match), according to a model, where the dependent variable $ch_{i,t}$ represents the subjects' choice (1 for Cooperate and 0 for Defect). We will estimate separately using the first part of each session (i.e. first 20 supergames) and then the second part. We will estimate using a logit estimator.

Let $p_{i,t}$ the probability of $ch_{i,t} = 1$ conditioned on the set of independent variables

$$p_{i,t} = \Lambda(\alpha_i + \beta[Ch_{i,t-1}; Partn.Ch_{i,t-1}] + \epsilon_{i,t}) \tag{1}$$

where $[Ch_{i,t-1}; Partn.Ch_{i,t-1}]$ is a 3-dimensional vector of dummy variables representing the different outcomes, where (1,0,0) represents $Ch_{i,t-1} = 0; Partn.Ch_{i,t-1} = 1$, (0,1,0) represents $Ch_{i,t-1} = 1; Partn.Ch_{i,t-1} = 0$ and (0,0,1) represents $Ch_{i,t-1} = 0; Partn.Ch_{i,t-1} = 0$; with mutual cooperation, $Ch_{i,t-1} = 1; Partn.Ch_{i,t-1} = 1$, being the baseline category. α_i is the time-invariant individual fixed-effect (taking into account time-invariant characteristics of both individuals and sessions); finally $\epsilon_{i,t}$ represent the error terms.

In table 9, we present the estimates of model 1 separately for the supergames belonging to the first and the second part of each sessions. Results are presented in odds ratios using the outcome $(C,C)_{t-1}$ as baseline. From Panel A, we note that the odd of cooperating at time t by high IQ are higher after that at least one in the match has defected (i.e. after $(D,D)_{t-1}$ $(D,C)_{t-1}$ $(C,D)_{t-1}$) when they play among themselves than when they play in the combined sessions. This difference is, if anything, even larger in the second part of the sessions as we can notice from Panel B of table 9.[16]. In table 10 we directly test whether high IQ are more forgiving when they play amongst each other than when play in combined sessions. We note that the high IQ are significantly less likely to cooperate whenever the other subject unilaterally defect. Interestingly, the low IQ do not seem to play systematically differently whether they play with with other low IQ or in the combined sections. Hence we can summarise this section by saying that *High IQ are less likely to cooperate after a unilateral deviation of the partner when they play in combined session than when play separately. The low IQ do not play in a systematically different way in the two treatments*

[16]for example, considering $(C,D)_{t-1}$, $0.03468 - 0.01485 > 0.01039 - 0.01450$

4 Beliefs' updating

There are essentially three ways cognitive abilities can affect the way subjects play: i) through more precise beliefs; ii) by best responding to their beliefs (in other words correctly calculating the expected payoffs of their choices); iii) and, after choosing a strategy, by being consistent with its implementation. We already saw that low IQ learn iii) when playing with high IQ and argued that this might be due to the fact that high IQ choose less forgiving strategies when matched with low IQ and this makes any mistake of the latter more costly. In what follows we will try to disentangle i) from ii) and understand better the mechanism by which low IQ learn from high IQ.

We assume that subjects in the first repeated game hold beliefs that other players either use AD or a cooperative strategy that we already defined SC (sophisticated cooperation = essentially corresponding to TfT + Grim). Closely following Dal Bó and Fréchette (2011), let the probability of player i in treatment group tr in supergame s to play AD be $\beta^{AD}_{i,tr,s}/(\beta^{AD}_{i,tr,s} + \beta^{SC}_{i,tr,s})$. In the first supergame, $s = 1$, subjects have beliefs charaterized by $\beta^{AD}_{i,tr,1}$ and $\beta^{SC}_{i,tr,1}$, from the second supergame onward, $s > 1$, the they update their beliefs as follows:

$$\beta^{k}_{i,tr,s+1} = \theta_{tr}\beta^{k}_{i,tr,s} + 1(a^{k}_{j,tr,s}), \tag{2}$$

where k is the action (AD or SC) and $1(a^{k}_{j,tr,s})$ takes the value 1 if the action of the partner j in treatment group tr in supergame s is k. The discounting factor of past belief, θ_{tr}, equals 0 in the so-called *Cournot Dynamics* and is 1 in *fictitious play.* Therefore as θ gets closer to 1, players react more strongly to overall partner past behaviour in updating their beliefs. Since we assume that subjects chose a strategy at the beginning of the supergame, they will play cooperation, C, in period 1 of supergame if they expect that the partner plays SC, defect, D, otherwise. The expected utility each player obtains for each action, a, is

$$U^{a}_{i,tr,s} = \frac{\beta^{AD}_{i,tr,s}}{\beta^{AD}_{i,tr,s} + \beta^{SC}_{i,tr,s}} u^{a}(a^{AD}_{j}) + \frac{\beta^{SC}_{i,tr,s}}{\beta^{AD}_{i,tr,s} + \beta^{SC}_{i,tr,s}} u^{a}(a^{SC}_{j}) + \lambda_{tr,s}\epsilon^{a}_{i,tr,s} \tag{3}$$

where $u^{a}(a^{k}_{j})$ is the payoff from taking action a when j takes the action k and $\lambda_{tr,s} = \lambda^{F}_{tr} + \phi^{s}\lambda^{V}_{tr}$ is the inverse of the capacity of best responding given the beliefs, which is allow to change over time.

In estimating the model we assume homogeneity of players within treatment groups in terms of the belief updating process. That is, individuals in the High-IQ split teatment group are assumed to have equal θ_{tr}, $\lambda_{tr,s}$, $\beta^{AD}_{i,tr,1}$ and $\beta^{SC}_{i,tr,1}$. The same assumption is made for individuals in Low-IQ split, High-IQ Combined and Low-IQ Combined treatment groups. Simulating choices according to the estimates obtained from this model generates choices of the first period of each supergame that on average fit well the experimental data as seen in figure 9.

We now analyse the two parameters we are interested in: θ_{tr}, measuring the strength of reaction to overall past partner behaviour while subjects update their beliefs and $\lambda_{tr,s}$, measuring the inverse of the capacity of best responding given the beliefs.[17]. The main parameters of the simulation are presented in table 11. We see that the θ_{tr} is higher for High-IQ split than Low-IQ split. This suggests that higher IQ subjects are putting more weight to overall past actions of their partners as compared to lower IQ subjects. Low-IQ split subjects instead appear to put more weight to the latest decision their partner has made. This difference dissipates in the combined treatment where we see that the estimated θ_{tr} is similar for both higher and lower IQ subjects. Furthermore, focusing on the $\lambda_{tr,s}$ estimates, we see that in both treatments (Split & Combined) higher IQ individuals are better able to best respond to their partners' actions given their beliefs.

[17] The details on how the model is estimated are in the online appendix of Dal Bó and Fréchette (2011) at p. 6-8.

5 Conclusions

In spite of the many forces operating in the direction of segregation of individuals along similarity of individual characteristics, a large part of social interaction occur across very diverse individuals. This occurs in particular across different levels of intelligence. So once it is clear that higher cognitive skills may favor a higher rate of cooperation, the natural question arises: what are the outcomes of strategic interactions among heterogeneous individuals. We have proved two main results.

The first is that cooperation rates in heterogeneous groups are close to the high cooperation rates, although the more intelligent makes a small loss. The entire aggregated surplus is higher when heterogeneous groups play than together than when they play separately, but the interaction in heterogeneous pooling is more advantageous to lower intelligence players.

The second result is that the higher cooperation rates of lower intelligence players in mixed groups is due to the influence of the choices of high intelligence players, who are more consistent in punishing defection when they play combined with less intelligent than when they play with subjects of a similar level of intelligence.

6 Figures and Tables

Figure 1: **Period 1 Cooperation for each supergame in all the Split and Combined sessions** Average cooperation over each supergame and session. High and all low IQ split treatments correspond to the the black and grey lines in the right panel, combined treatment correspond to the the middle grey lines in the left panel.

.

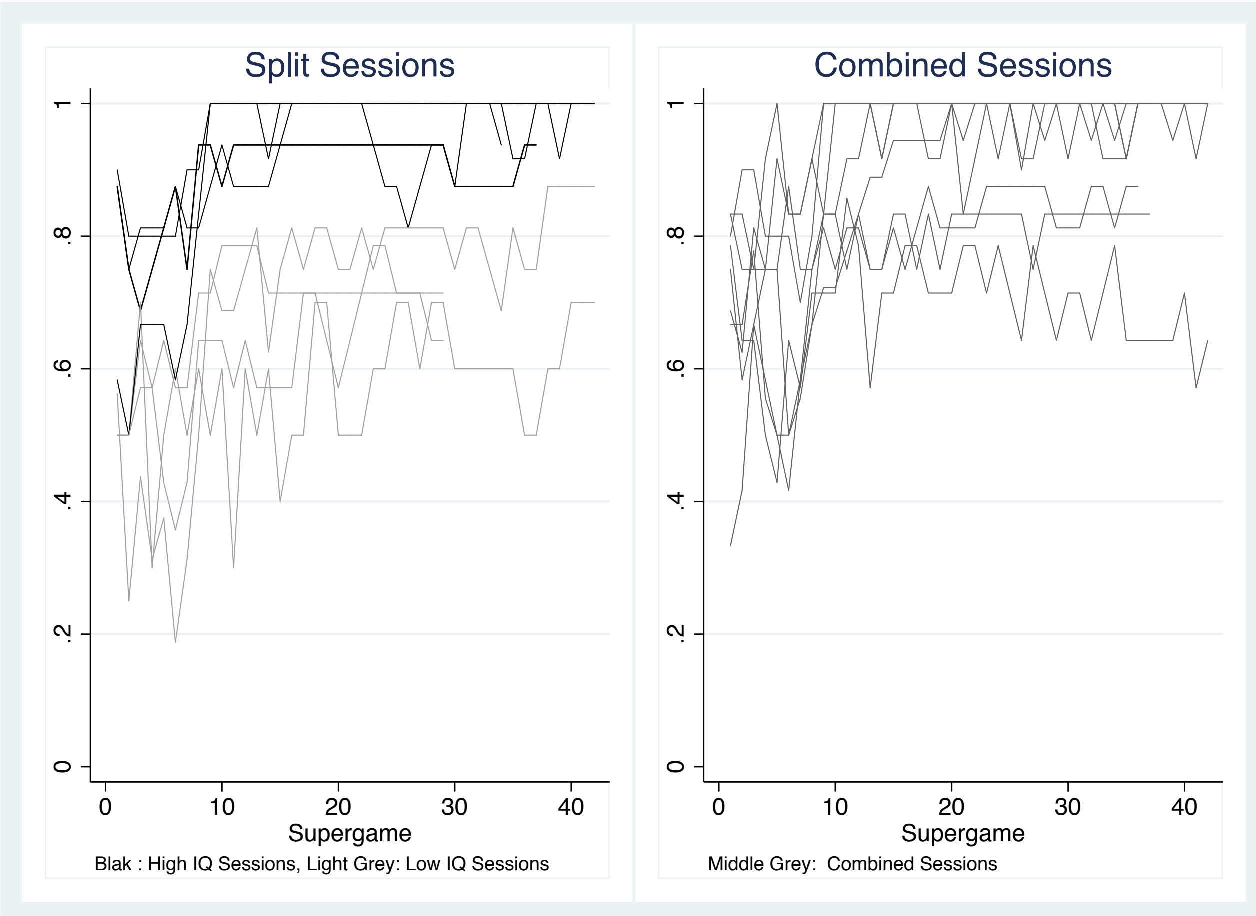
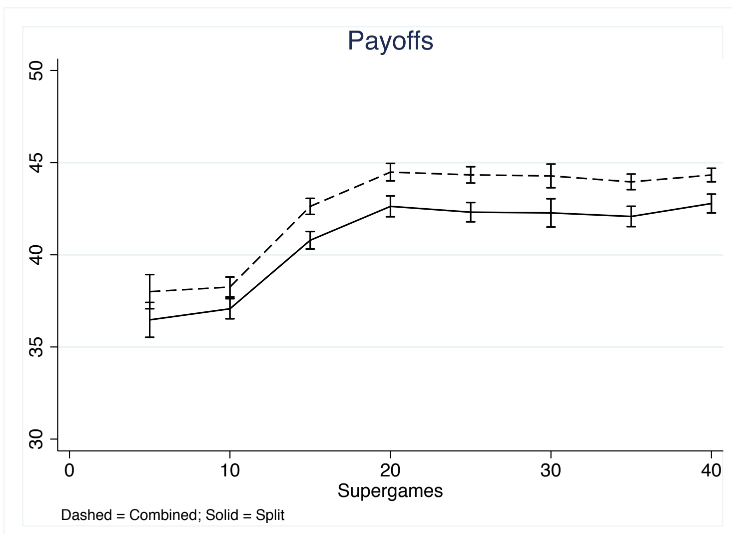


Figure 2: **Average payoffs per interaction in the Split and Combined sessions** The average is computed over observations in successive blocks of five supergames, of all Split and Combined sessions, aggregated separately. Bands represent 95% confidence intervals.



.

Figure 3: **Deviations from former cooperation over time.** A deviation from former cooperation is a choice of defect (D) at t following a round of mutual cooperation (C, C) at $t-1$. Bands represent 95% confidence intervals.

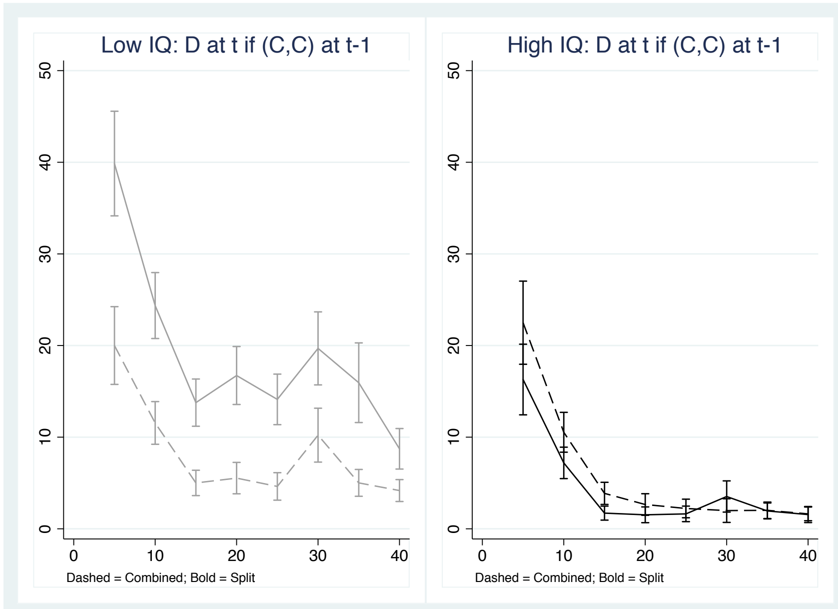


Figure 4: **Deviations from former cooperation across IQ classes.** Variability of the deviation from former cooperation in the two treatments, by quintile of IQ. Bands represent 95% confidence intervals.

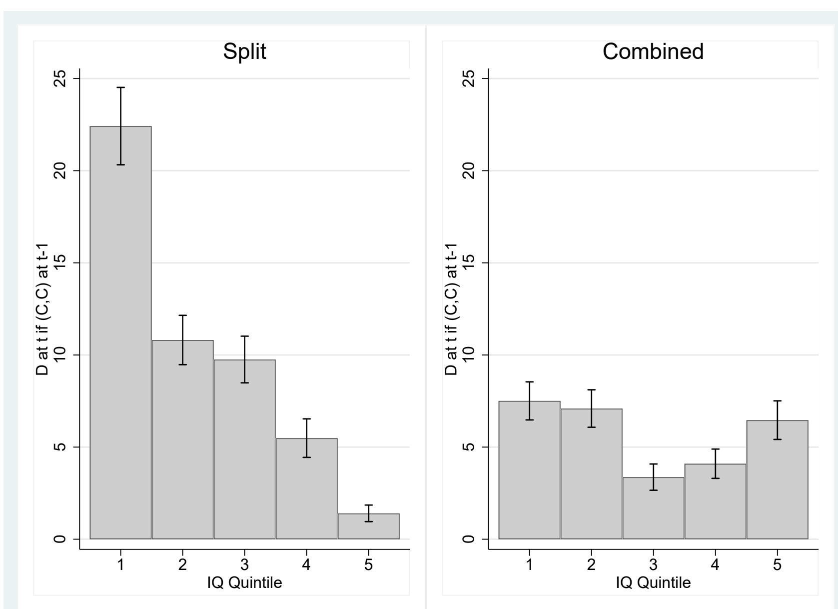


Figure 5: **Conditional Cooperation in the different groups and treatments.** Averages computed over observations aggregated by 5 supergames. Bands represent 95% confidence intervals

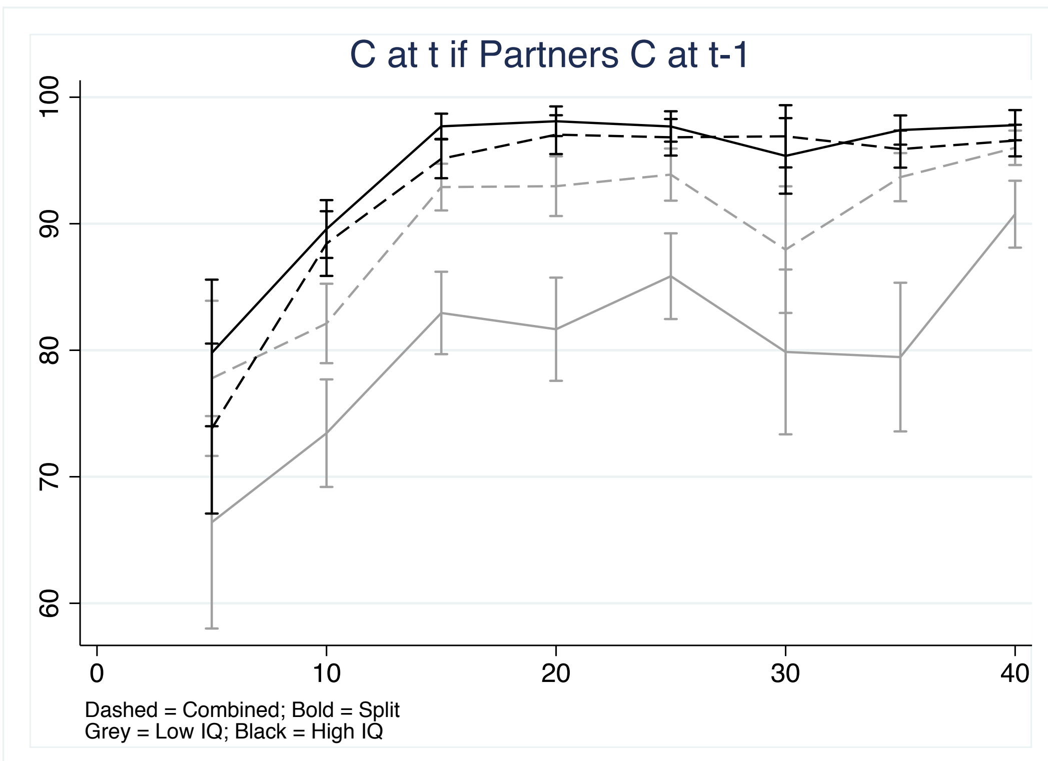


Figure 6: **Learning strategies for high-IQ subjects.** Averages computed over observations aggregated by supergames.

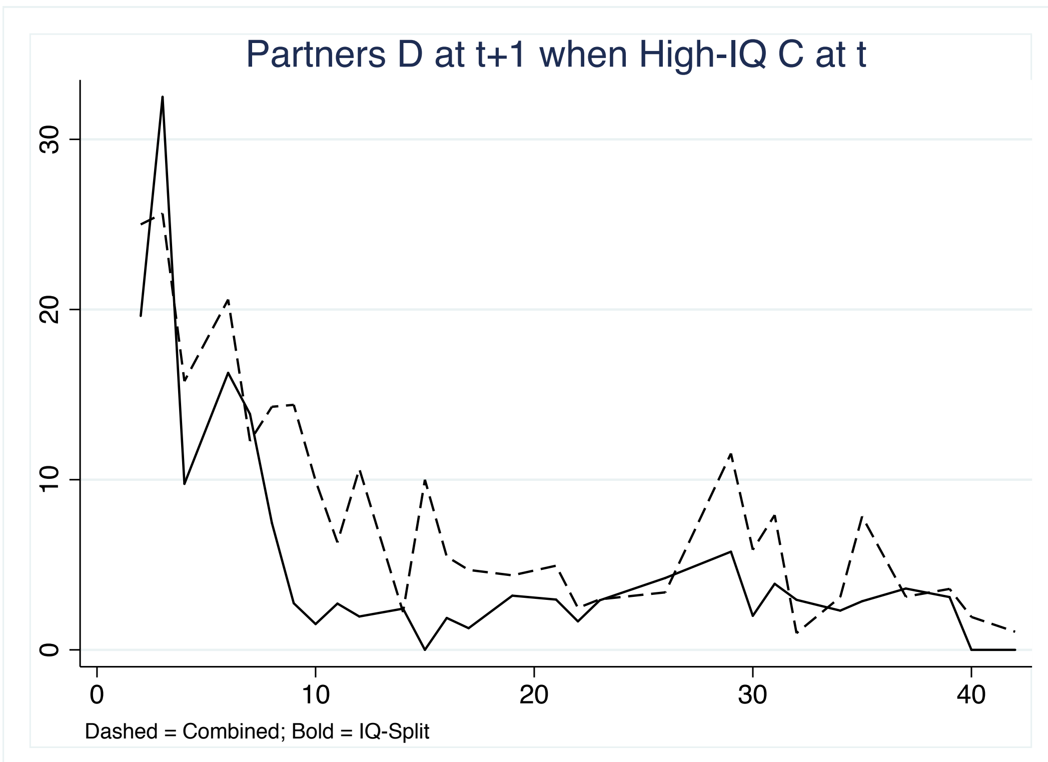


Figure 7: **Learning strategies for low-IQ subjects.** Averages computed over observations aggregated by supergames

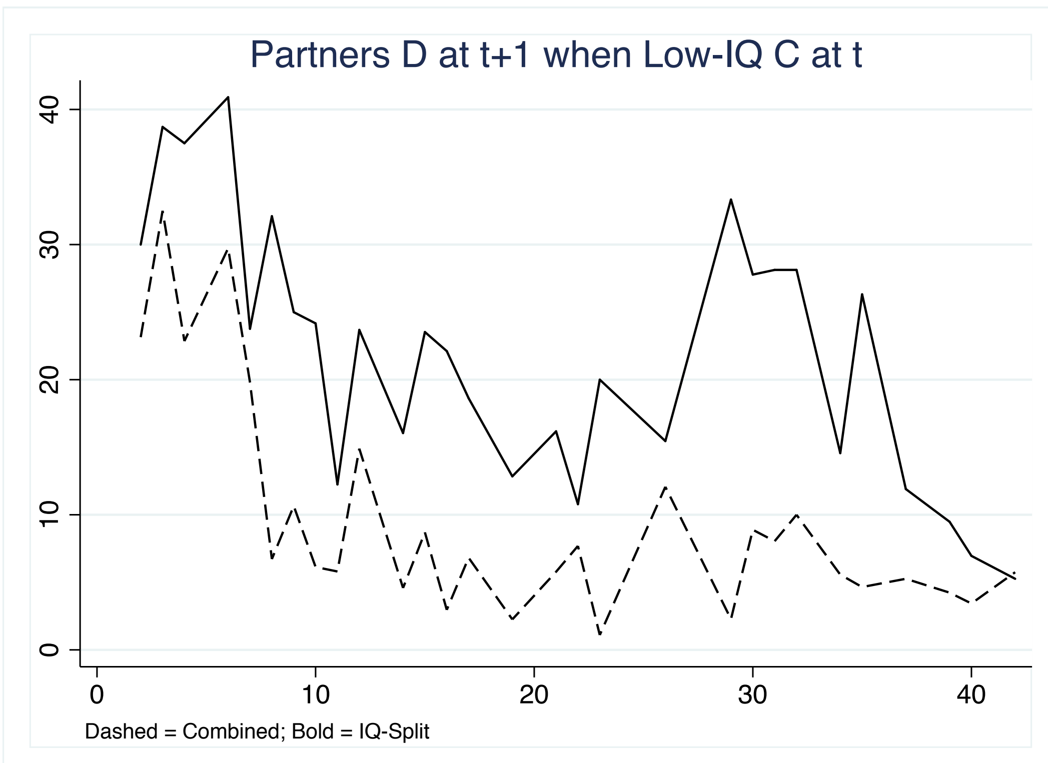


Figure 8: **Partners Defect at t+1 when High-IQ was suckered at t** The bars report average times partner chose D after an outcome (C, D) at period t. The vertical lines show the 5% confidence intervals

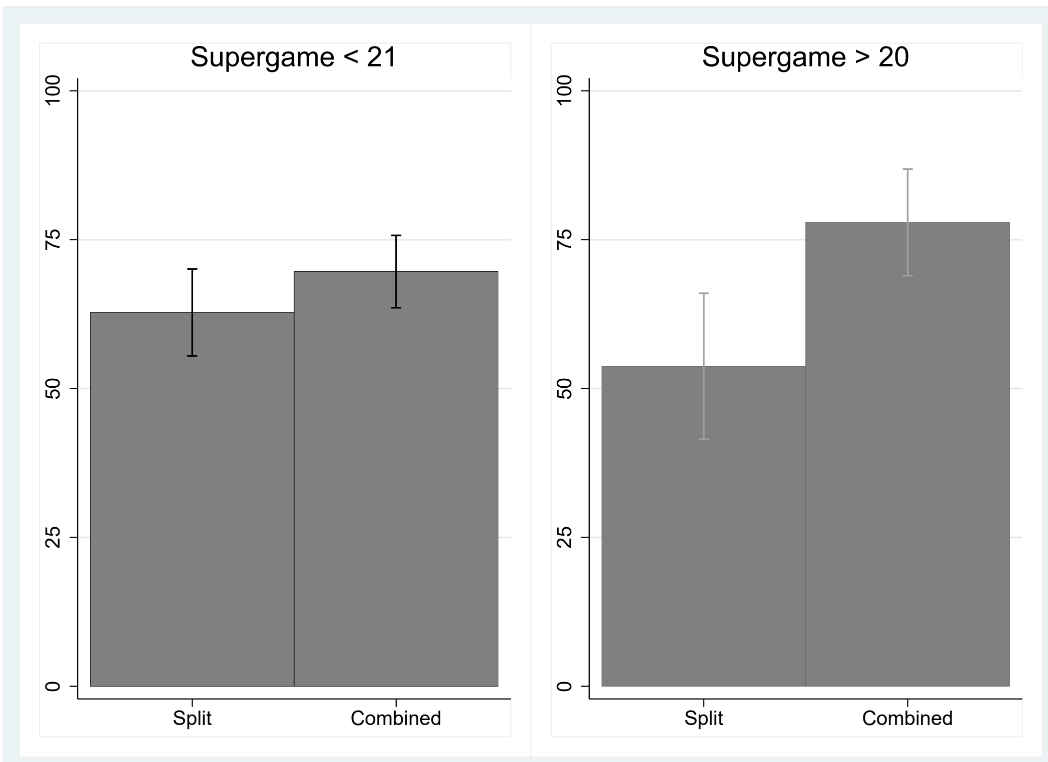


Figure 9: **Simulated Evolution of Cooperation Implied by the Learning Estimates** Solid lines represent experimental data, dashed lines the average simulated data, and dotted lines the 90 percent interval of simulated data.

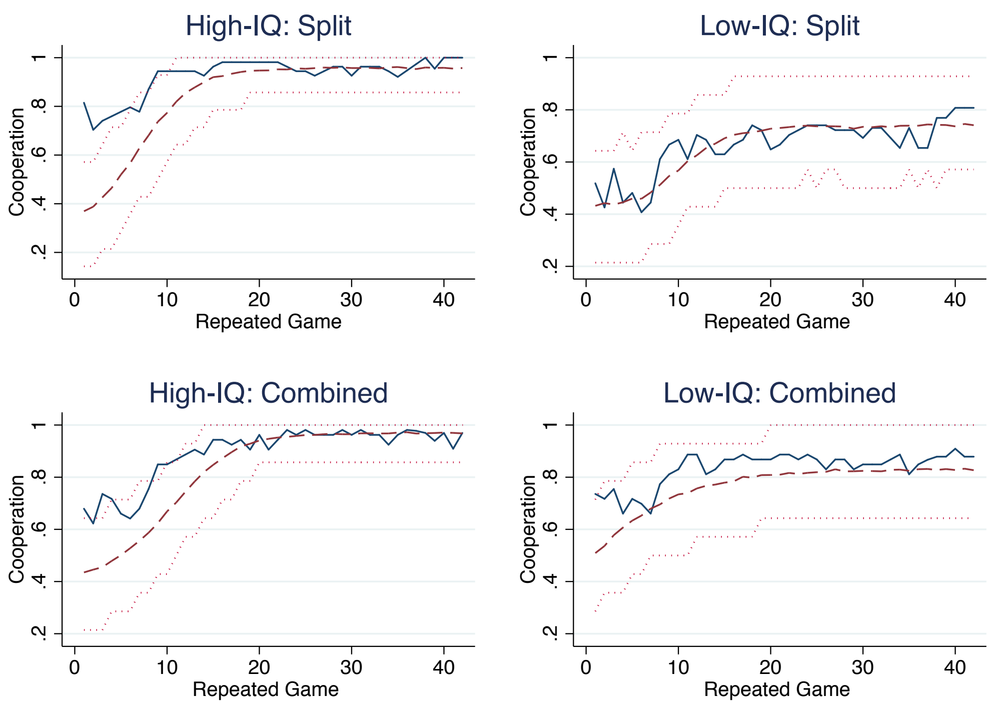


Table 2: **Effect of high IQ and low IQ session on choice of cooperation and payoffs** The dependent variables are average cooperation and average payoff across all interactions. The baseline are the combined sessions. OLS estimator. Robust standard errors clustered at the session levels in brackets; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$

	Supergame ≤ 20		Supergame > 20	
	Cooperate	Payoff	Cooperate	Payoff
	b/se	b/se	b/se	b/se
High IQ Session	0.0990**	2.5238**	0.0691	1.7259
	(0.0354)	(0.9217)	(0.0542)	(1.4115)
Low IQ Session	–0.2180***	–5.5977***	–0.2152***	–5.7067***
	(0.0524)	(1.3339)	(0.0612)	(1.5712)
# Subjects	–0.0112	–0.3063	–0.0062	–0.1812
	(0.0071)	(0.1815)	(0.0107)	(0.2766)
r2	0.203	0.407	0.152	0.320
N	214	214	214	214

Table 3: **Effects of IQ and other characteristics on cooperative choice in round 1 of each session** The dependent variable is the choice of cooperation in round 1. Logit estimator. **Note that coefficients are expressed in odds ratios**. Robust standard errors clustered at the session level; $p-values$ in brackets; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	Round 1 Cooperate b/p	Round 1 Cooperate b/p	Round 1 Cooperate b/p	Round 1 Cooperate b/p
choice				
IQ	1.00889		1.00942	
	(0.6444)		(0.6396)	
High IQ Group		1.76893		1.80835
		(0.1401)		(0.1358)
Extraversion			0.87817	0.91292
			(0.5544)	(0.6628)
Agreeableness			0.66879*	0.67223*
			(0.0681)	(0.0851)
Conscientiousness			1.21574	1.22401
			(0.4599)	(0.4356)
Neuroticism			0.75337	0.76481
			(0.3709)	(0.4035)
Openness			1.32202	1.32145
			(0.4504)	(0.4562)
Risk Aversion			0.79190***	0.79326***
			(0.0063)	(0.0095)
Age			0.99517	0.99802
			(0.9051)	(0.9605)
Female			1.04458	0.99468
			(0.8941)	(0.9872)
Combined Treatment			1.17291	1.16746
			(0.6737)	(0.6560)
Size Session			1.03245	1.02862
			(0.6398)	(0.6375)
N	214	214	214	214

Table 4: **Effects of split treatment on the evolution of cooperative choice in the first periods of all repeated games** The dependent variable is the choice of cooperation in the first periods of all repeated games. The baseline are the combined sessions. Logit with individual fixed effect estimator. Note that in the second part of each session many subjects made the same choices throughout, and for this reason their observations needed to be excluded from the estimations of the model in columns 3 and 4. Similar regressions with random effect (which does not need variability of choices at the individual levels avoiding this loss of observations) would deliver similar results. Std errors in brackets; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	Superg. ≤ 20		Superg. > 20	
	Cooperate	Cooperate	Cooperate	Cooperate
	b/se	b/se	b/se	b/se
choice				
High IQ Sessions*Supergame	0.14861***	0.15670***	–0.03499	0.01662
	(0.0502)	(0.0521)	(0.0666)	(0.0679)
Low IQ Sessions*Supergame	–0.06502**	–0.04342	0.08965**	0.09945**
	(0.0277)	(0.0285)	(0.0428)	(0.0456)
Supergame	0.12697***	0.09194***	–0.00911	–0.05359
	(0.0249)	(0.0257)	(0.0298)	(0.0372)
1st Per. Partners' Coop. at s-1		0.22917		1.16616***
		(0.1713)		(0.3479)
1st Per. Part. Coop. Rates until s-1		3.13168***		5.96293
		(0.5400)		(6.1902)
Partner Coop Rates until t-1		–0.24866		12.10323**
		(0.3303)		(5.0114)
Average lenght Supergame	0.69441***	0.78908***	1.74103**	1.79204**
	(0.1199)	(0.1312)	(0.8026)	(0.8556)
N	2280	2280	654	654

Table 5: **Split Treatment: Individual strategies in the different IQ sessions in the last 5 and first 5 subgames.** Each coefficient represents the probability estimated using ML of the corresponding strategy. Std error is reported in brackets. Gamma is the error coefficient that is estimated for the choice function used in the ML and beta is the probability estimated that the choice by a subject is equal to what the strategy prescribes.[(1)] Tests equality to 0 using the Wald test: * $p-values < 0.1$, ** $p-values < 0.05$ **, $p-values < 0.01$ ***

IQ Session	High	Low	High	Low
Repeated Games	Last 5	Last 5	First 5	First 5
Strategy				
Always Cooperate	0.3415**	0.1948**	0.1792**	0.1989**
	(0.1728)	(0.0850)	(0.0824)	(0.0991)
Always Defect	0.0302	0.2389***	0.1933***	0.4652***
	(0.0310)	(0.0668)	(0.0708)	(0.0915)
Grim after 1 D	0.3463	0.1855	0.2606**	0.0501
	(0.2295)	(0.1318)	(0.1203)	(0.0836)
Tit for Tat (C first)	0.2820	0.2599	0.3209**	0
	(0.2352)	(0.1800)	(0.1608)	(0.0830)
Win Stay Lose Shift	0	0	0.0460	0.0420
	(0.0326)	(0.0181)	(0.0890)	(0.0642)
Tit For Tat (after D C C)[(2)]	0	0.1207	0	0.2439**
Gamma	0.2794***	0.2911***	0.4662***	0.5578***
	(0.0495)	(0.0444)	(0.0468)	(0.0708)
beta	0.973	0.969	0.895	0.857
Average Rounds	4.82	4.52	1.8	1.8
N. Subjects	54	54	54	54
Observations	1,240	980	540	540

1. When beta is close to 1/2, choices are essentially random and when it is close to 1 then choices are almost perfectly predicted.
2. Tit for Tat (after D C C) stands for the Tit for Tat strategy that punishes after 1 defection but only returns to cooperation after observing cooperation twice from the partner.

Table 6: **Combined Treatment: Individual strategies in the different treatments in the last 5 and first 5 SGs.** Each coefficient represents the probability estimated using ML of the corresponding strategy. Std errors are reported in brackets. Gamma is the error coefficient that is estimated for the choice function used in the ML and beta is the probability estimated that the choice by a subject is equal to what the strategy prescribes.(1) Tests equality to 0 using the Waldtest: * $p-values < 0.1$, ** $p-values < 0.05$ **, $p-values < 0.01$ ***

IQ Partition	High	Low	High	Low
Repeated Games	Last 5	Last 5	First 5	First 5
Strategy				
Always Cooperate	0.0507	0.0545	0.0775	0.3158***
	(0.0807)	(0.0689)	(0.0765)	(0.1036)
Always Defect	0.0189	0.1321**	0.2623***	0.1858**
	(0.0288)	(0.0638)	(0.0703)	(0.0810)
Grim after 1 D	0.6289***	0.4580*	0.3949***	0.1564
	(0.1927)	(0.2392)	(0.1480)	(0.1064)
Tit for Tat (C first)	0.2452	0.3554	0.2654*	0.2068*
	(0.1618)	(0.2797)	(0.1378)	(0.1125)
Win Stay Lose Shift	0.0563	0	0	0.1353
	(0.0909)	(0.0064)	(0.0581)	(0.1166)
Tit For Tat (after D C C)(2)	0	0	0	0
Gamma	0.2353***	0.2722***	0.5270***	0.5236***
	(0.0263)	(0.0430)	(0.0655)	(0.0766)
beta	0.986	0.975	0.870	0.871
Average Rounds	5.12	5.12	1.8	1.8
N. Subjects	53	53	53	53
Observations	1,296	1,296	530	530

1. When beta is close to 1/2, choices are essentially random and when it is close to 1 then choices are almost perfectly predicted.
2. Tit for Tat (after D C C) stands for the Tit for Tat strategy that punishes after 1 defection but only returns to cooperation after observing cooperation twice from the partner.

Table 7: **IQ-Split: Payoffs at empirical frequency** The *Frequency* column reports the empirical frequency of each strategy in the set {*AC = Always Cooperate, AD = Always Defect, SC = Grim + TfT* } . The *Payoff* column reports the expected payoff using the strategy against the empirical frequency.

	High IQ		Low IQ		
	payoff	frequency	payoff	frequency	
AD	37.46	0.03	33.40	0.24	
AC	46.91	0.34	26.00	0.19	Last 5 Supergames
SC	47.49	0.63	43.99	0.57	

	High IQ		Low IQ		
	payoff	frequency	payoff	frequency	
AD	31.96	0.19	30.76	0.46	
AC	38.83	0.18	29.24	0.20	First 5 Supergames
SC	42.55	0.58	38.19	0.29	

Table 8: **Combined: Payoffs at empirical frequency** . The *Frequency* column reports the empirical frequency of each strategy in the set {*AC = Always Cooperate, AD = Always Defect, SC = Grim + TfT* } . The *Payoff* column reports the expected payoff using the strategy against the empirical frequency.

	High IQ		Low IQ		
	payoff	frequency	payoff	frequency	
AD	30.88	0.02	30.88	0.13	
AC	43.93	0.05	43.93	0.05	Last 5 Supergames
SC	45.38	0.87	45.38	0.81	

	High IQ		Low IQ		
	payoff	frequency	payoff	frequency	
AD	31.42	0.26	31.42	0.18	
AC	36.69	0.08	36.69	0.31	First 5 Supergames
SC	41.00	0.66	41.00	0.36	

Table 9: **Outcomes at period $t-1$ as determinants of cooperative choices at period t** The dependent variable is the cooperative choice at time t; the baseline outcome is mutual cooperation at $t-1$, $(C,C)_{t-1}$. Panel A relates to the first 20 supergames, panel B to the last 22 supergames. Logit with individual fixed effect estimator. **Coefficients are expressed in odds ratios;** $p-values$ **in brackets**; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

Panel A: $\#Supergame \leq 20$

	Low IQ Split	High IQ Split	Low IQ Combined	High IQ Combined
	b/p	b/p	b/p	b/p
choice				
$(C,D)_{t-1}$	0.00860***	0.01038***	0.00885***	0.00533***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$(D,C)_{t-1}$	0.01069***	0.01485***	0.00731***	0.01039***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$(D,D)_{t-1}$	0.00353***	0.00339***	0.00397***	0.00172***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N	2499	2448	2499	2448

Panel B: $\#Supergame > 20$

	Low IQ Split	High IQ Split	Low IQ Combined	High IQ Combined
	b/p	b/p	b/p	b/p
choice				
$(C,D)_{t-1}$	0.00301***	0.00527***	0.00426***	0.00153***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$(D,C)_{t-1}$	0.00402***	0.03468***	0.00270***	0.01450***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$(D,D)_{t-1}$	0.00121***	0.00318***	0.00157***	0.00044***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N	1718	1201	1771	1379

Table 10: **Outcomes at period $t-1$ as determinants of cooperative choices at period t** The dependent variable is the cooperative choice at time t; the baseline outcome is mutual cooperation at $t-1$, that is (C, C) at $t-1$. Combined is a dummy indicating the combined treatments. Logit with individual random effect estimator. Robust standard errors clustered at the session levels in brackets * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	High IQ	Low IQ
	All	All
	b/se	b/se
choice		
Combined*$(C,C)_{t-1}$	0.30868	0.39098
	(0.5137)	(0.3606)
Combined*$(D,D)_{t-1}$	–0.55593	0.32614
	(0.3414)	(0.4283)
Combined*$(D,C)_{t-1}$	–0.21615	–0.03074
	(0.2557)	(0.3078)
Combined*$(C,D)_{t-1}$	–0.52167**	0.38201
	(0.2580)	(0.3406)
$(D,D)_{t-1}$	–6.56678***	–6.41848***
	(0.4456)	(0.4022)
$(D,C)_{t-1}$	–4.69152***	–5.21715***
	(0.4560)	(0.2068)
$(C,D)_{t-1}$	–5.15376***	–5.27280***
	(0.2549)	(0.3545)
N	10343	10003

Table 11: Estimated Learning Parameters

Treatment Group	θ_{tr}	$\lambda_{tr,0}$	$\lambda_{tr,20}$	$\lambda_{tr,40}$
High-IQ: Split	.7180365	19.97237	19.97237	19.97237
Low-IQ: Split	.6212698	42.96482	42.96552	42.96483
High-IQ: Combined	.7742	18.25638	18.25638	18.25638
Low-IQ: Combined	.7977208	32.73199	32.82457	32.75003

Table 12: **Determinants of the response time** The dependent variable is the response time at period t; the baseline outcome are cooperative choice at t (for the choice of the player) and mutual cooperation at $t-1$, $(C,C)_{t-1}$ (for the history). (D_t) indicates the choice of defection by the player. Panel A relates to the first 20 supergames, panel B to the last 22 supergames. OLS with individual fixed effect estimator. Standard Errors in brackets; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

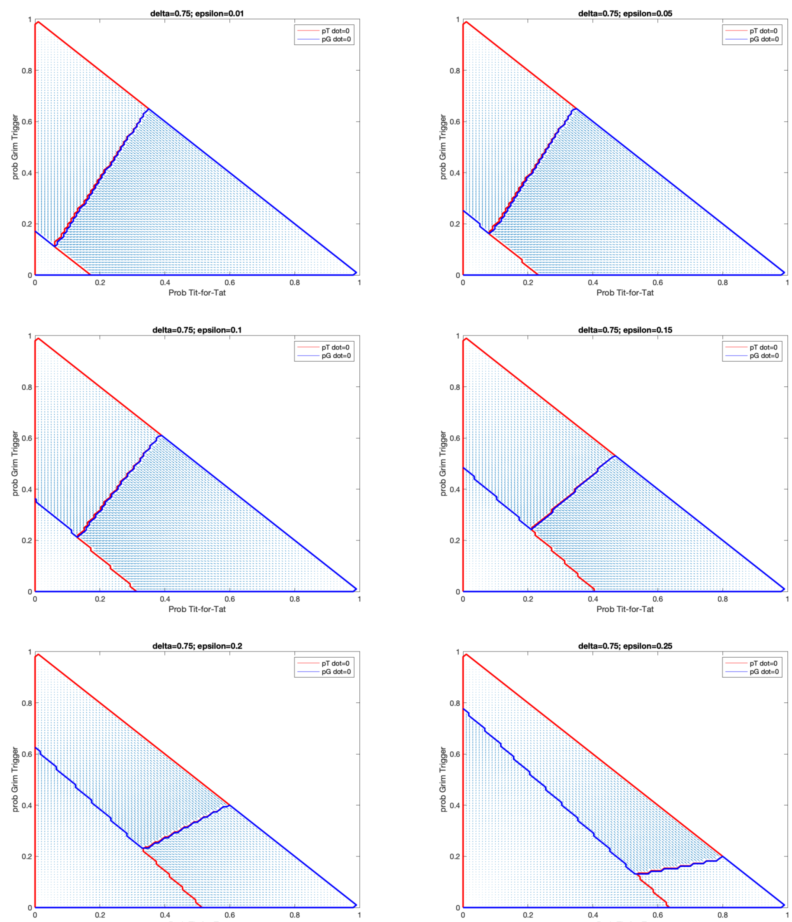
Panel A: $\#Supergame \leq 20$

	Low IQ IQ-Split b/se	High IQ IQ-Split b/se	Low IQ Combined b/se	High IQ Combined b/se
$(C,D)_{t-1}$	0.79349***	1.73748***	1.31364***	1.59534***
	(0.1957)	(0.2183)	(0.2216)	(0.2507)
$(D,C)_{t-1}$	0.66904***	2.52979***	1.09927***	2.13791***
	(0.2185)	(0.2292)	(0.2357)	(0.2717)
$(D,D)_{t-1}$	1.46417***	1.86581***	1.55534***	1.57749***
	(0.1670)	(0.1777)	(0.1742)	(0.2094)
N	2754	2754	2703	2703

Panel B: $\#Supergame > 20$

	Low IQ IQ-Split b/se	High IQ IQ-Split b/se	Low IQ Combined b/se	High IQ Combined b/se
$(C,D)_{t-1}$	0.04307	–0.13543	0.03826	–0.18413
	(0.1470)	(0.1753)	(0.1474)	(0.1410)
$(D,C)_{t-1}$	0.28546*	–0.20279	–0.04097	0.18021
	(0.1709)	(0.1925)	(0.1780)	(0.1476)
$(D,D)_{t-1}$	0.61658***	–0.02195	0.30837**	–0.15412
	(0.1148)	(0.1528)	(0.1232)	(0.1080)
N	1956	2296	2590	2590

Figure 10: **Basin of attraction of A, G and T, with transition error and Best Response dynamics.** The probability of error in transition is as displayed at the top of each panel, and is ranging from 1 per cent to 25 per cent. Payoff and discount factor ($\delta = 0.75$) are as in our experimental design.



A Design and Implementation: Additional Details

Table A.9 summarises the statistics about the Raven scores for each session in the IQ-split treatment and table A.10 for the Combined treatment. In the IQ-split treatment, the cutoff Raven score was 24 and 25. In sessions 7 and 8 the cutoff was 23 because the participants in these sessions scored lower on average than the rest of the participants in all the other sessions. Top-left panel of figure A.1 presents the overall distribution of IQ scores across both treatments. The bottom row of figure A.1 presents the distribution of the IQ scores across low- and high-IQ sessions for the IQ-split sessions, while top-right panel presents the distribution of the IQ scores for the Combined treatment sessions. Tables A.11 until A.13 present a description of the main data in the low- and high-IQ sessions in the IQ-split treatment and the Combined treatment sessions. Table A.14 shows the correlations among individual characteristics.

Table A.3 compares participant characteristics across the two treatments. Only the proportion of German participants is found to be significantly different across the two treatments, but as is obvious from tables A.4 and A.5 this is not significantly different across intelligence groups. Overall subjects are similiar across the two treatments. In table A.4 participant characteristics across intelligence groups in the IQ-split treatment are contrasted where only differences in the IQ scores are statistically different. Finally, table A.5 contrasts participant characteristics across intelligence groups across both treatments. As in table A.4 the only statistically significant difference is for IQ. Extraversion is found to be significantly different across intelligence groups but that cannot be reasonably seen as a driver of the results.

A timeline of the experiment is detailed below and all the instructions and any other pertinent documents are available online in the supplementary material.[18]

A.1 Timeline of the Experiment

Day One

1. Participants were assigned a number indicating session number and specific ID number. The specific ID number corresponded to a computer terminal in the lab. For example, the participant on computer number 13 in session 4 received the number: 4.13.

2. Participants sat at their corresponding computer terminals, which were in individual cubicles.

3. Instructions about the Raven task were read together with an explanation on how the task would be paid.

4. The Raven test was administered (36 matrices with a total of 30 minutes allowed). Three randomly chosen matrices out of 36 tables were paid at the rate of 1 Euro per correct answer.

5. The Holt-Laury task was explained verbally.

6. The Holt-Laury choice task was completed by the participants (10 lottery choices). One randomly chosen lottery out of 10 played out and paid

7. The questionnaire was presented and filled out by the participants.

[18] See note 3.

Between Day One and Two

1. Allocation to second day sessions made. An email was sent out to all participants listing their allocation according to the number they received before starting Day One.

Day Two

1. Participants arrived and were given a new ID corresponding to the ID they received in Day One. The new ID indicated their new computer terminal number at which they were sat.

2. The game that would be played was explained using en example screen on each participant's screen, as was the way the matching between partners, the continuation probability and how the payment would be made.

3. The infinitely repeated game was played. Each experimental unit earned corresponded to 0.004 GBP.

4. In the combined treatment participants completed a decoding task and a one-shot dictator game.

5. A de-briefing questionnaire was administered.

6. Calculation of payment was made and subjects were paid accordingly.

B Session Dates and Sizes

Tables A.1 and A.2 below illustrate the dates and timings of each session across both treatments.

Table A.1: **Dates and details for IQ-split**

Day 1: Group Allocation

	Date	Time	Subjects
1	23/04/2018	10:00	17
2	23/04/2018	11:00	19
	Total		36
3	07/05/2018	14:45	15
4	07/05/2018	16:00	11
	Total		26
5	12/06/2018	09:45	14
6	12/06/2018	11:30	19
	Total		33
7	20/11/2018	14:00	17
8	20/11/2018	15:15	19
	Total		36

Day 2: Cooperation Task

	Date	Time	Subjects	Group
Session 1	25/04/2018	10:00	16	High IQ
Session 2	25/04/2018	11:30	14	Low IQ
	Total Returned		30	
Session 3	09/05/2018	14:00	10	High IQ
Session 4	09/05/2018	15:30	10	Low IQ
	Total Returned		20	
Session 5	14/06/2018	10:00	12	High IQ
Session 6	14/06/2018	11:30	14	Low IQ
	Total Returned		26	
Session 7	22/11/2018	14:00	16	High IQ
Session 8	22/11/2018	15:30	16	Low IQ
	Total Returned		32	
	Total Participants		108	

Table A.2: **Dates and details for Combined**

Day 1: Group Allocation

	Date	Time	Subjects
1	30/04/2018	09:45	7
2	30/04/2018	11:00	13
	Total		20
3	15/05/2018	10:00	6
4	15/05/2018	11:30	16
	Total		22
5	18/06/2018	14:45	17
6	18/06/2018	16:00	9
	Total		26
7	10/07/2018	09:45	7
8	10/07/2018	11:00	13
	Total		20
9	02/10/2018	09:45	7
10	02/10/2018	11:00	11
	Total		18
11	15/10/2018	09:45	6
12	15/10/2018	11:00	6
	Total		12

Day 2: Cooperation Task

	Date	Time	Subjects
Session 1	02/05/2018	10:00	14
Session 2	17/05/2018	14:00	10
Session 3	17/05/2018	15:30	12
Session 4	20/06/2018	14:00	12
Session 5	20/06/2018	15:30	12
Session 6	12/07/2018	10:00	18
Session 7	04/10/2018	11:30	16
Session 8	17/10/2018	11:30	12
	Total Participants		106

Table A.3: **Comparing Variables across the IQ-Split and the Combined Sessions**

Variable	Split	Combined	Differences	Std. Dev.	N
IQ	103.4069	103.1394	.2674614	1.349413	214
Age	23.84259	23.06604	.7765549	.6392821	214
Female	.4907407	.5	-.0092593	.0686773	214
Openness	3.767593	3.678302	.0892907	.0730968	214
Conscientiousness	3.358025	3.431866	-.0738411	.0883303	214
Extraversion	3.228009	3.371462	-.143453	.1024118	214
Agreeableness	3.591564	3.612159	-.0205955	.0850711	214
Neuroticism	3.016204	2.879717	.1364867	.0995567	214
Risk Aversion	5.536082	5.382979	.1531038	.251421	191
German	.6481481	.754717	-.1065688	.0624657**	214
Total Profit	5167.87	5957.415	-789.5447	141.8649***	214
Rounds Played	126.8519	139.8302	-12.97834	2.591088***	214
Payoff per Round	40.19059	41.89426	-1.703675	.6099137***	214
Total Profit (Equal SGs Played)	3858.296	4021.849	-163.5528	57.84501**	214
Payoff per Round (Equal SGs Played)	40.19059	41.89426	-1.703675	.6025522**	214

Note: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$

Table A.4: **Comparing Variables across IQ-split Sessions**

Variable	Low IQ	High IQ	Differences	Std. Dev.	N
IQ	95.94193	110.8718	-14.92987	1.232502***	108
Age	24.14815	23.53704	.6111111	1.142875	108
Female	.462963	.5185185	-.0555556	.0969619	108
Openness	3.824074	3.711111	.112963	.0975451	108
Conscientiousness	3.376543	3.339506	.037037	.1160422	108
Extraversion	3.386574	3.069444	.3171296	.1456155**	108
Agreeableness	3.609054	3.574074	.0349794	.1201571	108
Neuroticism	2.949074	3.083333	-.1342593	.1357823	108
Risk Aversion	5.652174	5.431373	.2208014	.394149	97
German	.6111111	.6851852	-.0740741	.0924877	108
Final Profit	4481.481	5854.259	-1372.778	184.8242***	108
Rounds Played	122.4815	131.2222	-8.740741	4.266736**	108
Payoff per Round	36.68508	44.50096	-7.815882	.5747042***	108
Total Profit (Equal SGs Played)	3480.667	4235.926	-755.2593	55.6599***	108
Payoff per Round (Equal SGs Played)	36.25694	44.12423	-7.867284	.5797906***	108

Note: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$

Table A.5: **Comparing Variables across IQ-split Groups Across both Treatment Sessions**

Variable	Low IQ	High IQ	Differences	Std. Dev.	N
IQ	95.68959	110.8592	-15.1696	.8576931***	214
Age	23.83178	23.08411	.7476636	.6394164	214
Female	.4672897	.5233645	-.0560748	.0685692	214
Openness	3.741122	3.705607	.035514	.0733099	214
Conscientiousness	3.425753	3.363448	.0623053	.0883684	214
Extraversion	3.398364	3.199766	.1985981	.1019719**	214
Agreeableness	3.613707	3.589823	.0238837	.0850633	214
Neuroticism	2.925234	2.971963	-.046729	.0999411	214
Risk Aversion	5.451613	5.469388	-.0177749	.2517194	191
German	.7102804	.6915888	.0186916	.0628772	214
Final Profit	5177.28	5940.626	-763.3458	142.5326***	214
Rounds Played	131.0748	135.486	-4.411215	2.723199*	214
Payoff per Round	39.30087	43.82866	-4.527786	.5416761***	214
Total Profit (Equal SGs Played)	3729.673	4148.944	-419.271	51.40749***	214
Payoff per Round (Equal SGs Played)	38.85076	43.21817	-4.367407	.5354947***	214

Note: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$

Figure A.1: **Distribution of IQ Scores.** Top-left panel shows IQ distribution for all participants across both treatments, top-right shows IQ distribution in Combined treatment and bottom panels show IQ distribution in low- and high-IQ sessions from IQ-split treatment.

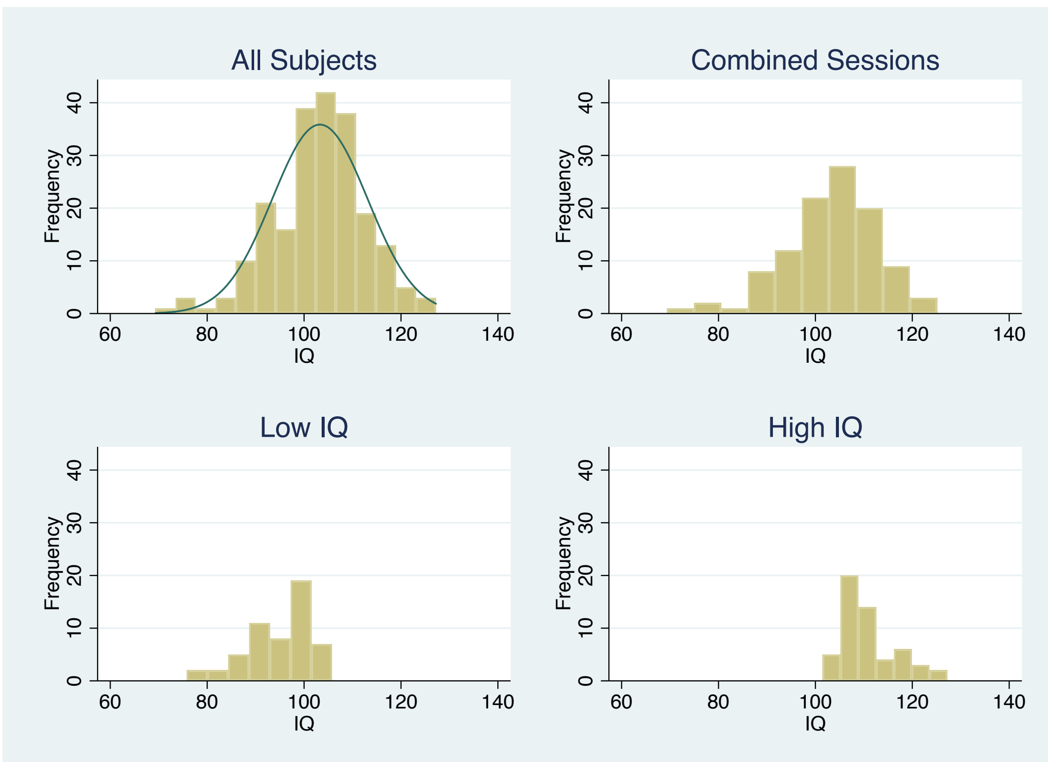


Table A.6: **Countries of Origin of Participants**

Country	Number	Percentage
Albania	2	0.93
Belarus	1	0.47
Bulgaria	2	0.93
Canada	1	0.47
China	9	4.21
Denmark	1	0.47
Egypt	3	1.40
France	1	0.47
Germany	150	70.09
Hungary	1	0.47
India	3	1.40
Indonesia	1	0.47
Italy	4	1.87
Japan	1	0.47
Kazhakhstan	1	0.47
Kosovo	1	0.47
Moldova	2	0.93
Peru	1	0.47
Poland	1	0.47
Romania	1	0.47
Russia	7	3.27
Serbia	1	0.47
Spain	3	1.40
Switzerland	2	0.93
Syria	1	0.47
Taiwan	1	0.47
Turkey	4	1.87
UK	1	0.47
USA	2	0.93
Ukraine	4	1.87
Vietnam	1	0.47
Total	214	100.00

Table A.7: **SGs and Rounds Played by Session in IQ-Split**

Session	SGs	Rounds
1	37	123
2	29	96
3	42	151
4	42	151
5	40	146
6	29	96
7	34	116
8	42	151

Table A.8: **SGs and Rounds Played by Session in Combined**

Session	SGs	Rounds
1	42	151
2	42	151
3	42	151
4	37	123
5	42	151
6	42	151
7	36	119
8	37	123

Table A.9: Raven Scores by Sessions in IQ-split Treatment

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
High IQ - Session 1	28.063	2.886	25	35	16
Low IQ - Session 2	20.214	3.725	11	24	14
High IQ - Session 3	28	2.539	25	33	10
Low IQ - Session 4	22	2.539	18	25	10
High IQ - Session 5	27.917	3.147	24	34	12
Low IQ - Session 6	19.357	3.671	11	23	14
High IQ - Session 7	25.875	2.029	23	31	16
Low IQ - Session 8	20.5	2.394	15	23	16

Table A.10: Raven Scores by Sessions in Combined Treatment

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Session 1	23.214	5.754	11	31	14
Session 2	22	6.532	8	31	10
Session 3	22.833	4.859	13	31	12
Session 4	25.333	3.339	20	32	12
Session 5	24.917	2.466	20	29	12
Session 6	24.833	4.19	16	32	18
Session 7	23.375	4.674	16	30	16
Session 8	23	4.533	16	34	12

Table A.11: IQ-split: Low IQ Sessions, Main Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Choice	0.561	0.496	0	1	6614
Partner Choice	0.561	0.496	0	1	6614
Age	23.983	7.876	17	65	6614
Female	0.454	0.498	0	1	6614
Round	64.824	40.281	1	151	6614
Openness	3.85	0.518	2.5	5	6614
Conscientiousness	3.37	0.559	2.333	4.667	6614
Extraversion	3.408	0.683	1.875	4.75	6614
Agreableness	3.585	0.680	1.667	4.889	6614
Neuroticism	2.969	0.696	1.125	5	6614
Raven	20.552	3.04	11	25	6614
Risk Aversion	5.639	2.016	0	10	6614
Final Profit	4695.723	1037.735	3168	6337	6614
Profit x Period	36.685	3.179	28.669	42.875	54
Total Periods	122.481	27.739	96	151	54

Table A.12: IQ-split: High IQ Sessions, Main Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Choice	0.875	0.331	0	1	7086
Partner Choice	0.875	0.331	0	1	7086
Age	23.518	3.28	18	33	7086
Female	0.523	0.5	0	1	7086
Round	66.91	39.264	1	151	7086
Openness	3.723	0.497	2.6	4.8	7086
Conscientiousness	3.322	0.64	1.444	4.556	7086
Extraversion	3.073	0.816	1.25	4.625	7086
Agreableness	3.578	0.563	2	5	7086
Neuroticism	3.081	0.707	1.375	4.375	7086
Raven	27.44	2.745	23	35	7086
Risk Aversion	5.386	1.63	2	9	7086
Final Profit	5941.262	864.996	4312	7248	7086
Profit x Period	44.501	2.78	36.382	48	54
Total Periods	131.222	14.615	116	151	54

Table A.13: Combined, Main Variables

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Choice	0.795	0.404	0	1	14822
Partner Choice	0.795	0.404	0	1	14822
Age	23.048	2.906	18	33	14822
Female	0.496	0.5	0	1	14822
Round	71.156	41.578	1	151	14822
Openness	3.683	0.553	2.4	5	14822
Conscientiousness	3.432	0.684	1.556	4.778	14822
Extraversion	3.378	0.73	1.625	4.625	14822
Agreableness	3.614	0.61	2.111	4.889	14822
Neuroticism	2.872	0.743	1.375	4.625	14822
Raven	23.759	4.621	8	34	14822
Risk Aversion	5.407	1.508	2	9	14822
Final Profit	6026.931	851.060	3984	7212	14822
Profit x Period	42.555	3.933	30.417	47.762	106
Total Periods	139.83	14.467	119	151	106

Table A.14: All participants: Correlations Table (p−*values* in brackets)

Variables	Raven	Female	Risk Aversion	Openness	Conscientiousness	Extraversion	Agreableness	Neuroticism
Raven	1.000							
Female	0.003	1.000						
	(0.969)							
Risk Aversion	0.018	0.148	1.000					
	(0.795)	(0.030)						
Openness	-0.025	0.016	-0.059	1.000				
	(0.715)	(0.814)	(0.390)					
Conscientiousness	-0.043	0.070	0.093	0.026	1.000			
	(0.533)	(0.307)	(0.175)	(0.707)				
Extraversion	-0.151	0.080	0.015	0.298	0.202	1.000		
	(0.028)	(0.243)	(0.831)	(0.000)	(0.003)			
Agreableness	-0.024	0.181	0.055	0.068	0.287	0.170	1.000	
	(0.732)	(0.008)	(0.423)	(0.324)	(0.000)	(0.013)		
Neuroticism	0.041	0.287	-0.002	0.092	-0.192	-0.276	-0.112	1.000
	(0.551)	(0.000)	(0.972)	(0.178)	(0.005)	(0.000)	(0.102)	

From: XXX <XXX@gmail.com>
Sent: Thursday, November 22, 2018 5:53 PM
To: andis.sofianos@uni-heidelberg.de
Subject: Re: REMINDER: Allocations for second part of Decision Task Experiment

Hi,

This is a feedback about this interesting experiment I've got the chance to take part into today at 15h30 ! I hope that it will not compromize the anonymity of the experiment, otherwise just do not pay attention to this email :)

I don't know if you are the one who has made it up, but I just wanted to say to you that I've found this experiment very stimulating, in so far as I unterstood its principle (so, which was the better way to cooperate with the partner in order to obtain the maximal benefit for both of us) maybe just 5 minutes before the end. Well, I could say that I was tired, but it wouldn't be fair to make up an excuse... :)

Indeed at the beginning I just paid attention to my own gain table to see where I could ear the maximal benefit, and so I was always picking up the letter D in the hope to get 50. That was always working at the beginning with one of my partner, so I though that I had taken the most sensible decision indeed, and kept up with my D choice. But with an other partner (and then, with the first partner, too), I always won 50 at the beginning, as my partner was choosing the letter C for our first round, and thus maximising my gain, but after my first answer was always changing his decision to D, so that I could only win 25. If then I tried to come back to C, he was still blocking me by keeping on chosing D (and thus making me "lose", so that I could only earn 12), so that I had to follow the "tempo" he was giving, because I didn't wanted to try to make him change his mind to C again into choosing C several times myself, because I was afraid that it could take time before he would understand that I was ready to cooperate again, and during this time it was always better for me to earn 25 rather than 12.
But then I understood that his way to always choose C at the beginning was a way to "test" me, to see if I was willing to cooperate in order that we could both of us earn 48 : if I had "selfishly" chosen D during this first round (thus earning my 50 but making him "lose" with just 12), he was "revenging" by getting the 50 for him and letting me earn only 25, using my fear to be then left stranded with only 12 (which was not so important for him, because he was then at least getting 25) to make me follow his rule.
After a (too long !) moment, I finally get it, and the cooperation perfectly worked in so far as I was accepting his "invitation" and choosing the letter C from the beginning of the partnership, thus showing my good will : he then kept on to the letter C too, so that up to the moment of my "revelation" we've always made 48, both of us.
For my defence, my primary choice to take D at the first round could have been sensible, but only if I could have known before that it would only be a "one shot" with the partner, thus taking advantage on his trust into my will to cooperate to got the 50 and let him with the 12, and "swapping partner" just after...

So, thank you for matching me which such clever and adaptative partners, and I would be very interested into reading more about this experiment of yours/ its results (and about this prisoner game you spoke about and that I didn't know), and into taking part in more similar experiments, which I've found very stimulating ! In the future, would it be possible to register only for experiments of this kind, and not to get any more offers of "stress experiments" about mental calculation, that I do not find so interesting than this king of decision tasks ?

Faithfully yours,

XXXXX

C Alternative formal analysis of beliefs' updating

In our analysis of the data we showed that less intelligent subjects play less efficiently in the split treatments and learn how to play more efficiently when mixed with more intelligent. There are essentially three ways cognitive abilities can affect the way subjects play: i) through more precise beliefs; ii) by best responding to their beliefs (in other words correctly calculating the expected payoffs of their choices); iii) and, after choosing a strategy, by being consistent with its implementation. We already saw that low IQ learn iii) when playing with high IQ and argued that this might be due to the fact that high IQ choose less forgiving strategies when matched with low IQ and this makes any mistake of the latter more costly. In what follows we will try to disentangle i) from ii) and understand better the mechanism by which low IQ learn from high IQ.

We assume that subjects in the first repeated game hold beliefs that other players either use AD or a cooperative strategy that we already defined SC (sophisticated cooperation = essentially corresponding to TfT + Grim). Closely following Dal Bó and Fréchette (2011), let the probability of player i in supergame s to play AD be $\beta^{AD}_{i,s}/(\beta^{AD}_{i,s}+\beta^{SC}_{i,s})$. In the first supergame, $s=1$, subjects have beliefs charaterized by $\beta^{AD}_{i,s}$ and $\beta^{SC}_{i,1}$, from the second supergame onward, $s>1$, the they update their beliefs as follows:

$$\beta^{k}_{i,s+1}=\theta_i\beta^{k}_{i,s}+1(a^k_j), \tag{A-1}$$

where k is the action (AD or SC) and $1(a^k_j)$ takes the value 1 if the action of the partner j is k. The discounting factor of past belief, θ_i, equals 0 in the so-calleld *Cournot Dynamics* and is 1 in the *fictitious play.* Therefore the closer is θ to 1 the slower will player update their beliefs. Since we assume that subjects chose a strategy at the beginning of the supergame, they will play cooperation, C, in period 1 of supergame if they expect that the partner plays SC, defect, D, otherwise. The expected utility each player obtains for each action, a, is

$$U^a_{i,s}=\frac{\beta^{AD}_{i,s}}{\beta^{AD}_{i,s}+\beta^{SC}_{i,s}}u^a(a^{AD}_j)+\frac{\beta^{SC}_{i,s}}{\beta^{AD}_{i,s}+\beta^{SC}_{i,s}}u^a(a^{SC}_j)+\lambda_{i,s}\epsilon^a_{i,s} \tag{A-2}$$

where $u^a(a^k_j)$ is the payoff from taking action a when j takes the action k. The estimation of the model above generates choices of the first period of each supergame that in average fits well our data as it is shown in figure A.2. We now analyse the two parameters we are interested: θ_i, measuring the inverse of the speed by which subjects update their beliefs and $\lambda_{i,s}$, measuring the inverse of the capacity of best responding given the beliefs.[19].

In table A.17, we show the correlation between IQ and the parameters of interest. IQ significantly negatively correlated with θ_i, implying that higher IQ subjects update faster their beliefs. While do not affect the capacity of best responding, λ_s. In the top panels of figure A.3 we can compare the cumulative distribution of the θ_i in the different treatments. θ_i seem to be smaller for high IQ than for low IQ, confirming that low IQ update they beliefs slower than high IQ (top left panel). When combined the differences seem to be drastically reduced (top right panel). From panel A of table A.16 , we note that the differences between high IQ and low IQ in the split treatment is statistically significant, while the same difference in the combined treatment is only weakly significant at the best. The bottom left panel of figure A.3 shows that low IQ improve their speed (i.e. θ_i is lower) when combined with the high IQ, while there is no much difference among high IQ subjects in the different treatments. Panel B of A.16 confirm that the differences among the low IQ in the combined and in the split treatments are statistically significant. We can summarise this

[19]The details on how the model is estimated are in the online appendix of Dal Bó and Fréchette (2011) at p. 6-8., the main parameter of the simulation are presented in table A.15 of the appendix

discussion saying *Less intelligent learn to update their beliefs faster when they are mixed with more intelligent, while the way the subjects best respond to their beliefs is not depended on their IQs.* A possible explanation of why lower IQ subjects update their first period beliefs faster when mixed with the higher IQ might be that in the latter environment they receive a clearer signal from the other players playing more consistent strategies of cooperation.

Figure A.2: **Simulated Evolution of Cooperation Implied by the Learning Estimates** Solid lines represent experimental data, dashed lines the average simulated data, and dotted lines the 90 percent interval of simulated data.

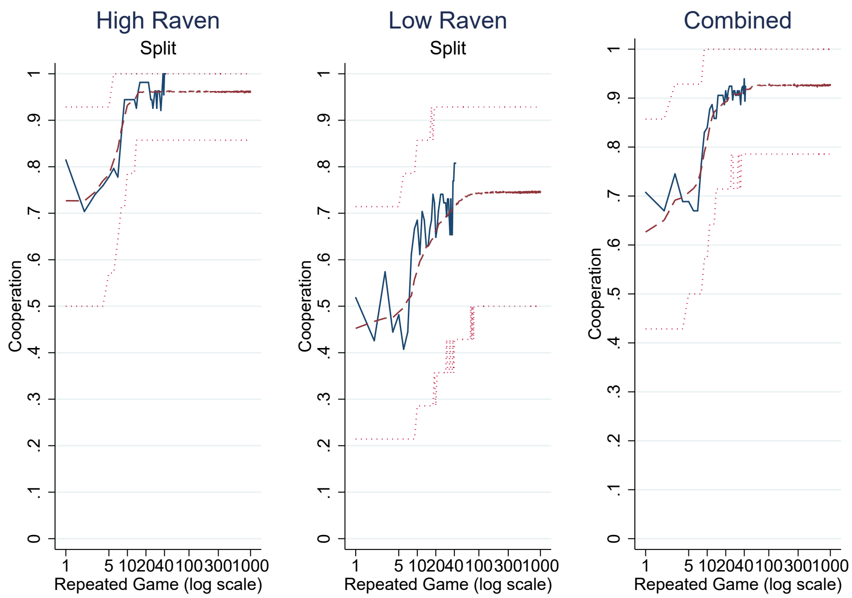


Figure A.3: **Distribution of the beliefs' updating speed within the different groups and treatments**. Distribution of the parameter θ_i as defined in equation A-1, where 1 correspond to slowest speed (fictitious play) and 0 to the fasted speed (Cournot dynamics)

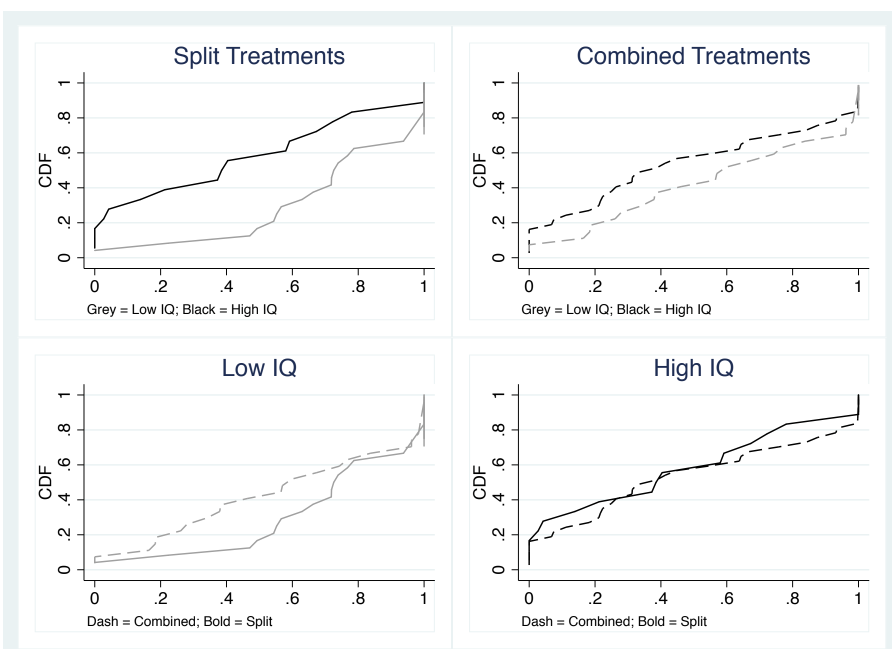


Table A.15: IQ and Simulated Parameters

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
IQ	103.516	10.203	69.338	127.231	182
θ_i	0.58	0.357	0	1	129
λ_0	5.67	11.683	0	93.275	129
λ_{20}	7.28	12.941	0.154	93.275	128
λ_{40}	6.846	12.913	0.141	93.275	128

Table A.16: **Differences in the beliefs' updating speed within the different groups and treatments**. Tests of the differences of the estimated parameter θ_i as defined in equation A-1, where 1 correspond to slowest speed (fictitious play) and 0 to the fasted speed (Cournot dynamics) * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

Panel A: Tests between IQ groups

Treatment		Split	Combined
		θ_{LowIQ} - θ_{HighIQ}	θ_{LowIQ} - θ_{HighIQ}
t test	t	-2.9623^{***}	-1.3777^{*}
Mann-Witney	z	-2.488^{**}	-1.411

Panel B: Tests between treatments

Treatment		Split vs Combined	Split vs Combined
		θ_{LowIQ}	θ_{HighIQ}
t-test	t	1.9647^{**}	-0.3909
Mann-Witney	z	1.849^{*}	-0.350

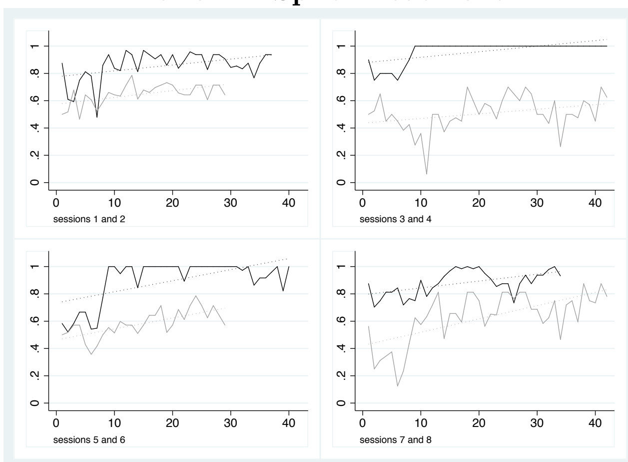
Table A.17: **Correlation between IQ, beliefs updating and capacity of best responding to own beliefs** Correlations between IQ, updating speed, θ_i (as defined in equation 2), capacity of best responding to beliefs in supergame s, λ_s (as defined in equation 3). $p-values$ in brackets

Variables	IQ	θ_i	λ_0	λ_{20}	λ_{40}
IQ	1.000				
θ_i	-0.345	1.000			
	(0.000)				
λ_0	-0.032	-0.242	1.000		
	(0.746)	(0.006)			
λ_{20}	-0.047	-0.196	0.887	1.000	
	(0.635)	(0.026)	(0.000)		
λ_{40}	0.001	-0.205	0.899	0.988	1.000
	(0.992)	(0.020)	(0.000)	(0.000)	

C.1 Supplementary Data Analysis

Figure A.4: **Average cooperation per supergame in all different sessions** The grey lines in each panel represent the average cooperation per period among all subjects of the corresponding low IQ groups and the black lines represent the average cooperation per supergame among all subjects of the corresponding high IQ groups. The dashed lines represent the combined sessions, the bold lines the split sessions, and the dotted straight lines the linear trends.

Panel A: Split Treatment



Panel B: Combined Treatment

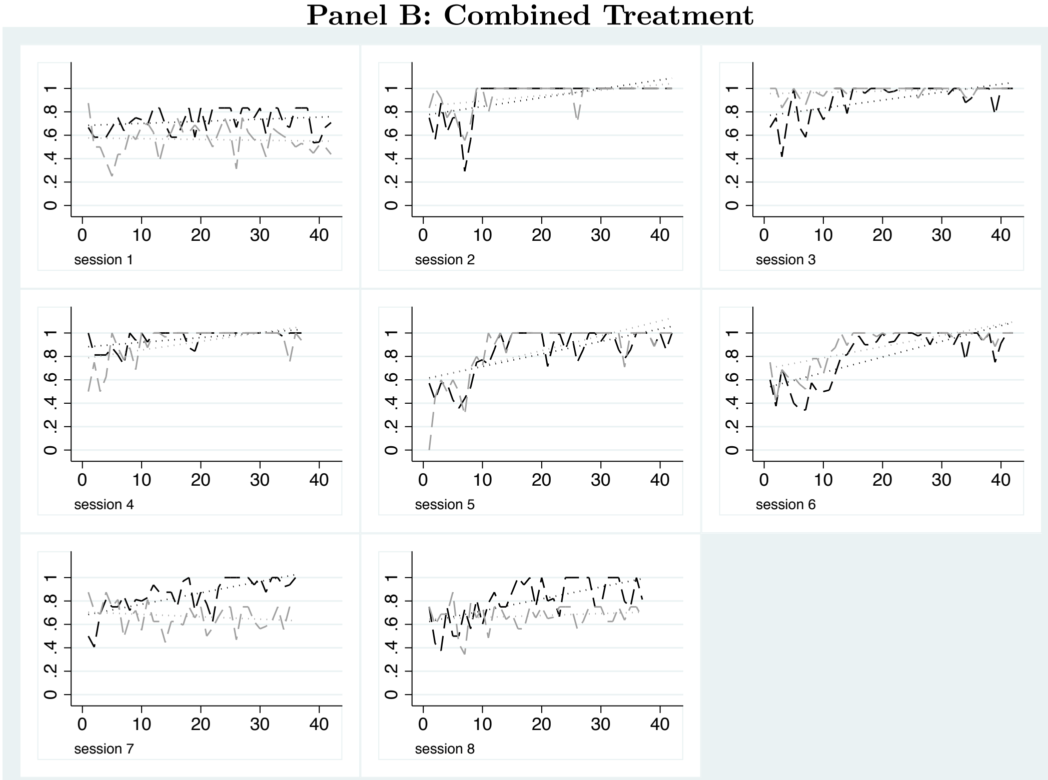


Table A.18: **Determinants of the response time in periods 1 of each supergame** The dependent variable is the response time in period 1. (D_t) indicates the choice of defection by the player. Combined sessions is the baseline in column 1. OLS estimator with random effect. Std errors clustered at the individual levels in brackets; * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	Periods 1 R. Time b/se	Periods 1 R. Time b/se	Periods 1 R. Time b/se	Periods 1 R. Time b/se
(D_t)	0.66084***	0.64697***	0.53450	0.65654***
	(0.2173)	(0.2170)	(2.0029)	(0.2191)
High IQ Session	0.02999			
	(0.1790)			
Low IQ Session	–0.05748			
	(0.1710)			
IQ		–0.01150**	–0.01172**	–0.01088*
		(0.0058)	(0.0052)	(0.0064)
IQ*(D_t)			0.00111	
			(0.0190)	
Extraversion	–0.12382	–0.15458*	–0.15495*	–0.14663*
	(0.0849)	(0.0826)	(0.0821)	(0.0770)
Agreeableness	0.13042	0.13081	0.13106	0.19555*
	(0.1170)	(0.1179)	(0.1177)	(0.1089)
Conscientiousness	0.04954	0.05130	0.05127	–0.03450
	(0.0994)	(0.0979)	(0.0978)	(0.0926)
Neuroticism	–0.03161	–0.03375	–0.03396	–0.09047
	(0.0806)	(0.0833)	(0.0832)	(0.0811)
Openness	–0.10066	–0.09467	–0.09424	0.00687
	(0.1345)	(0.1310)	(0.1298)	(0.1203)
Risk Aversion	–0.06234	–0.06259	–0.06269	–0.05126
	(0.0461)	(0.0452)	(0.0452)	(0.0433)
Age	–0.00579	–0.00915	–0.00915	–0.01204
	(0.0159)	(0.0152)	(0.0152)	(0.0131)
Female	–0.11406	–0.10646	–0.10635	–0.06894
	(0.1453)	(0.1412)	(0.1411)	(0.1241)
# Supergame	–0.03156***	–0.03174***	–0.03174***	–0.03087***
	(0.0047)	(0.0047)	(0.0047)	(0.0047)
Size Session	0.00551	0.00597	0.00594	
	(0.0237)	(0.0237)	(0.0236)	
Partner C rates until t-1	–0.89242***	–0.84457**	–0.84218***	–0.93145***
	(0.3444)	(0.3303)	(0.3254)	(0.3512)
Average Supergame Length	–1.30962***	–1.31495***	–1.31488***	–1.31030***
	(0.0967)	(0.0963)	(0.0964)	(0.0972)
Sessions Fixed-Effects	No	No	No	Yes
N	7961	7961	7961	7961

D A Model of Errors and Strategy Evolution: Analysis

D.1 Errors in Action Choice and in Transition

It is clear that subjects can make two types of errors: the error in the complex process of observing the action of the others, recalling the rules of the game and their plan of action, and deciding what to in the future, that we have represented here (by modeling subjects' plans as choice of automata) by a choice of the next period state in their automaton. In line with this modeling, we can consider the possibility that they also make an error in their choice of action when they are at a state in the automaton. This possibility turns out to produce only a minor change in the analysis of the evolution of the frequency of actions. We therefore focus in the main text on the error in the transition, which is closest to the important difference among subjects.

However, errors in action choice have of course to be allowed in our data analysis. So we analyze this possibility, although we report the results in the appendix (section (C))

D.2 Why stricter strategies thrive with larger errors

The intuitive reason for the results is the following. When no error is possible, the strategies G and T produce the same outcome when matched with A and when matched with each other. So with no error both equilibria and dynamic behavior are indeterminate.

When errors are possible, the first crucial fact is (see Lemma (G.3)) that the A strategy is for every error probability the unique best response to itself. Hence, no matter what the error is, the profile (A, A) stays a Nash equilibrium and a locally stable steady state. If some strategy is eliminated, that is not going to be A. The second crucial fact is that at $\epsilon = 1/2$ under some condition on payoffs (which is satisfied by the payoffs in the experimental design we adopted) the (A, A) is the unique Nash (see Lemma (G.4)). These two facts together with the continuity of the value function in ϵ (see Lemma (G.1)) imply that the set of strategies that survive in a learning-selection model change from the entire set to the strict strategy A. What is left to study is what happens during the transition.

When a small error is possible, the value for (say) player 1 of using G with G (call it $V^1(GG)$) compared to the similarly defined values $V^1(TG)$, $V^1(GT)$, $V^1(TT)$ are no longer equal. Small changes of the values are sufficient to break the indeterminacy, and in fact we will see that locally unique, locally stable equilibria emerge.

Two forces are now in place. To clarify them we consider as example how the share of the frequency (the relative size of the basin of attraction) between G and T changes. As we mentioned already, in the model with with no errors there is meaning for basin of attraction. For infinitesimally small errors, the value of the steady state that splits the unit interval into the two regions of attraction is determined by the ratio of the derivatives of the gains in values for the strategy profiles (see the sub-section H.1). However the gains (coming from small changes in error) for T are larger than those for G, hence the basin of attraction of T is larger for small errors. This is first force, that pins down the relative fraction of the two strategies, and establishes (in the limit) the benchmark for the "almost no error" model. However, there is a second force: as the error becomes larger, the difference in gains becomes smaller, and thus the fraction of T declines. If we consider the limit case of a fifty-fifty probability of error, one can see that the only surviving strategy is the strictest strategy, A.

D.3 The Role of the Discount Factor

The effect of the discount factor on the basin of attraction follows naturally from the effects of δ on payoffs. Let us consider first the case of δ large, close to 1. Consider players starting at the mutual cooperation. A small transition error can lead one to a defection. In this case the long run loss induced by a small error in the transition of the G strategy is large: with grim trigger, the state can go back to a cooperative state only by another mistake. Instead, the effects of such an error with Tit-for-Tat are not as durable. Hence TfT increases fitness, and its basin of attraction is larger.

We have seen the comparative static effect of an increase in δ. The increase in the error rate (modeled by an error probability ϵ) reduces the comparative advantage of the lenient strategy. In the limit case of an equal probability of error and correct choice, the difference between the two strategies disappears. Thus, as the error probability increases, the basin of attraction of Tit-for-Tat decreases as compared to that of grim trigger (the stricter strategy). This is exactly what we see in the data. Our data analysis suggests that the pathway of the effect is the one we identify in the model.

D.4 Evolution and Learning

The intuitive arguments we have provided so far to explain the reason for our main result that a higher probability of error leads to a wider use of strict strategies are independent of the specific evolutionary model adopted. The reason for this is that what dictates the relative fitness of strategies is of course ultimately their relative profitability, and so the effect of the frequency of errors is its effect on the payoff structure. We will use to argue this point two different models: one is the replicator dynamics, or (if we want to emphasize the social learning model behind the specific functional form of the change in frequency), the Proportional Imitation Dynamics model (PID). The other is the Best Response Dynamics model (BRD).

E Setup

We are now ready for a more detailed and precise treatment of our problem. We consider a Repeated Prisoner's Dilemma game with stage game payoff u: where

$$\begin{array}{ccc} & C^2 & D^2 \\ C^1 & c, c & s, t \\ D^1 & t, s & d, d \end{array}$$

$$t > c, d > s, t + s \leq 2c. \tag{A-3}$$

Note that these conditions are satisfied by the payoffs in our experimental design, with a strict inequality for the last.

We focus on the set of repeated game strategies $\{A, G, T\}$, where A is the *Always Defect* strategy, G is the *Grim Trigger*, and T is the *Tit-for-Tat*.

We consider in the following the automaton representation of these three strategies for each player (not for the pair of players), with states D_A for A, C_G and D_G for G, and C_T and D_T for T. When we refer to an automaton we may omit the index of the player who is using that automaton, relying on the symmetry of the game. Here an automaton M is a tuple (X, x_0, f, P) where X is the set of states of the automaton, x_0 the initial state, f is a function $X \to A^i$, where A^i is the set of actions of player i. Finally, P defines the transition probability where

$$P(\cdot; x, (a^1, a^2)) \in \Delta(X) \tag{A-4}$$

We prefer the notation in terms of transition probability rather than functions (in spite of the fact that the transition are deterministic) to allow a smooth transition to the later case in which we introduce errors.

In the analysis later, when we introduce errors in transition, we will want to consider the automaton G_C and G_D, having the same transition and same action choice function as G, but having C and D as initial state; so G_D will be different from A because the state of the automaton G_D may transit back to C by mistake, whereas the state of A can never transit to a state where C is chosen. T_C and T_D are defined similarly.

F The Equilibrium and evolutionary dynamics with no transition errors

We consider the normal form game where the strategy set for each player is the set

$$M \equiv \{A, G, T\} \tag{A-5}$$

Players chooses simultaneously an element in M, and the payoff is the one induced in the repeated game by the pair of strategies or automata, reported in **??**:

	A	G	T
A	d, d	$(1-\delta)t + \delta d, \delta d$	$(1-\delta)t + \delta d, \delta d$
G	$\delta d, (1-\delta)t + \delta d$	c, c	c, c
T	$\delta d, (1-\delta)t + \delta d$	c, c	c, c

We call this normal form game induced by the choice of strategies the *strategy choice* (SC) game. This game is special, in that the two "actions" G and T are interchangeable. The analysis of the game with three actions can be reduced to the analysis of the *reduced* game with two actions $\{A, GT\}$ with payoffs:

	A	GT
A	d, d	$(1-\delta)t + \delta d, \delta d$
GT	$\delta d, (1-\delta)t + \delta d$	c, c

We denote by μ_R the strategies in the reduced game.

When $c < (1-\delta)t + \delta d$ then A is a dominant strategy, so there is a unique equilibrium of the reduced game (and thus of the original strategy choice game) at (A, A). Multiple equilibria are possible when δ is larger than the critical value

$$\delta^* \equiv \frac{t-c}{t-d}. \tag{A-6}$$

We consider in the following only the case in which

$$\delta > \delta^* \tag{A-7}$$

in this case there are three equilibria, with the mixed strategy equilibrium assigning a probability to A given by:

$$\mu_R(A,\delta)^* = \frac{c-(1-\delta)t-\delta d}{c-(1-\delta)t-\delta d+(1-\delta)d} \tag{A-8}$$

Proposition F.1. *When (A-7) holds, the reduced game has two equilibria in pure strategies (A, A) and (GT, GT)) and a mixed strategy equilibrium, with $\mu_R(A,\delta)^*$ the probability of A. For any such equilibrium there is a corresponding continuum of equilibria in the strategy choice game, where the probability $\mu_R(GT,\delta)^*$ is assigned arbitrarily to the strategies G and T.*

Corresponding to these equilibria there is a set of steady states in the evolutionary dynamics we consider now.

F.1 Evolutionary Dynamic

We let $\mu \in \Delta(M)$ denote a mixed strategy and also a frequency of choice of strategy in the population (Sandholm (2007), Weibull (1997)). When we consider the evolution over time of the frequency, we let $\mu(t,\cdot)$ denote the vale of the frequency in the population at t. We denote the payoff to a player adopting m when the frequency in the population is μ as $u(m,\mu)$, and for any $\tau \in \Delta(M)$,

$$u(\tau,\mu) \equiv \sum_{m\in M} \tau(m)u(m,\mu)$$

The time evolution of the frequency under proportional imitation is

$$\forall m \in M, \frac{d\mu(t,m)}{dt} = \mu(t,m)\left(u(m,\mu)-u(\mu,\mu)\right) \tag{A-9}$$

The best response correspondence is defined as usual:

$$BR(\mu) \equiv \{\tau \in \Delta(M) : \forall m \in M, u(\tau,\mu) \geq u(m,\mu)\} \tag{A-10}$$

The time evolution of the frequency under best response dynamics is described by the differential inclusion:

$$\forall m \in M : \frac{d\mu(t,m)}{dt} \in BR(\mu(t,\cdot))(m) - \mu(t,m). \tag{A-11}$$

F.2 Evolutionary Dynamic with no Error

We indicate mixed strategies or frequencies in the following as $(\mu(A), \mu(G), \mu(T))$.

The dynamic behavior under proportional imitation has a natural long run behavior: when the proportion of players choosing A is large enough, only A survives in the long run; conversely when the initial fraction of A is sufficiently low, only cooperative strategies (G and T) survive. In this second case, however, the long run relative weight of G and T is entirely determined by the initial conditions, and is not robust: a small change in the initial conditions alters the long run behavior. More precisely:

Proposition F.2. *Under both proportional imitation and best response dynamics the following hold:*

1. *the set of steady states consists of the singleton* $(1,0,0)$*; the interval*

$$\{(\mu^*, \mu^* p, \mu^*(1-p)) : p \in [0,1] \tag{A-12}$$

and the interval

$$\{(0, p, (1-p)) : p \in [0,1]\}; \tag{A-13}$$

2. the path with initial condition $\mu(0, A) > \mu^*$ converge to the steady state $(1, 0, 0)$; the paths with initial conditions where $\mu(A) < \mu^*$ converge to a steady state in the set (A-13) above;

3. The steady states in (A-12) are all unstable.

In addition, in the proportional imitation dynamic the paths with initial condition $(\mu(0, A), \mu(0, G), \mu(0, T))$ are straight-lines where for every time t:

$$\frac{\mu(t, G)}{\mu(t, T)} = \frac{\mu(0, G)}{\mu(0, T)}.$$

Thus a path with initial condition $\mu(0, A) > \mu^*$ converges to the steady state

$$\left(0, \frac{\mu(0, G)}{\mu(0, G) + \mu(0, T)}, \frac{\mu(0, T)}{\mu(0, G) + \mu(0, T)}\right)$$

Evolutionary dynamic in the case we are considering (automata with no memory error) cannot select among possible relative frequencies of G and T, thus cannot address the issue of whether a more lenient strategy (such as T) is more or less frequent in the long run than the stricter strategy G. The long run relative frequency is whatever frequency happened to be there in the initial condition. We now show that the frequency is precisely determined in the long run when an error of arbitrarily small size is possible.

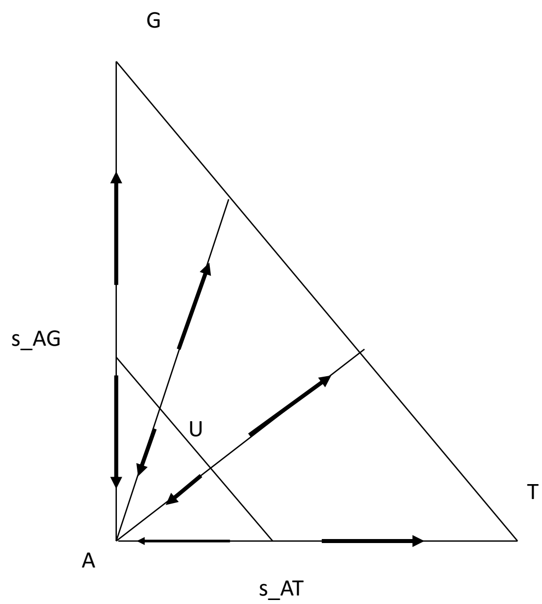
Figure (A.5) below illustrates the dynamic behavior of the proportional imitation model when no errors occur. The dynamic behavior with the Best Response dynamic is very similar, but in such degenerate case (in which the payoffs for the strategies G and T are same) there is a large multiplicity of solutions of the differential inclusion; so we prefer the proportional imitation dynamic for purposes of illustration.

The dynamics is drawn on the projection of the simplex representing the frequency of the three strategies A, G and T on the triangular region $(\mu(G), \mu(T))$, so that the vertex denoted by the letters correspond to the pure strategy denoted by that letter. The vertex A indicates the only locally unique, locally stable equilibrium, with basin of attraction the entire triangular region below the shorter segment (labelled U) in the interior of the triangular region. Any point on the line joining G and T is an equilibrium of the strategy choice game and a steady state. Similarly all the points on U are unstable states.

G The Value Function with transition errors

We now consider the case which is relevant for our experimental data. Subject can choose a strategy, that we have described as an automaton (a set of rules they have to follow), but they have to implement the transition relying on their memory of the relevant bits of information: the current state, the action profile, the transition rules. In implementing the rule they may transit to the wrong state. We allow them to forget some element that determines the next state, but not the automaton they are using (e.g. the may forget what the other did, or what the state was, but they do not forget what automaton they were using.) In our case, the set of states for each automaton consists of two elements, so the error can only take the form of choosing the other (wrong) state. The probability of this error is $\epsilon > 0$. The link with our experimental data is provided by such parameter, which describes the an individual characteristic: the higher the Intelligence, the lower the ϵ.

Figure A.5: **Proportional Imitation dynamics with no error.** All points on the segment joining s_{AG} and s_{AT} (denoted by U) are unstable steady states. All points the segment joining G and T are stable steady states. All lines joining A with points on the latter segment are invariant. Only few are illustrated here.



G.1 Value function equation

To compute the payoff from the choice of the strategy when transition error are possible we need to extend the state space to include explicitly the automata that are produced by errors, distinguishing two automata on the basis of their internal state. These sets were only appearing implicitly in our previous analysis. The extended strategy set, S, is:

$$S \equiv \{A, G_c, G_d, T_c, T_d\} \tag{A-14}$$

where G_c is the G (Grim Trigger) automaton with C as initial state, (D for G_d). There is a unique action determined by an automaton in S, and will be denoted by $a(s)$.

The payoff from a pair of choices of initial automata made by the two players is determined by a simple recursive equation on functions defined on the product space Ω:

$$\Omega \equiv S \times S \tag{A-15}$$

with generic element $\omega = (s^1, s^2)$.

G.1.1 Transitions

Note that with this notation we can write the transition for the automaton G in state C to the same automaton in state D as the transition from G_c to G_d. So we can define the transition on the set S (keeping the notation P) with $P(s'; s, a^1, a^2)$ the probability of transiting to s' if the current state is s and the action profile is (a^1, a^2). [20]

[20] More precisely, we can write this transition from the point of view of player 1:

We now turn to the definition of the transition function with errors. We let $\mathcal{Q}$ the set of stochastic matrices on Ω. First, we let the transition with no errors to be denoted by Q, where

$$\forall \omega = (s^1, s^2), \omega' = (r^1, r^2), \text{ if } a(\omega) \equiv (a^1(s^1), a^2(s^2)),$$
$$Q(\omega'; \omega) \equiv P(r^1; s^1, a(\omega))P(r^2; s^2, a(\omega)) \tag{A-16}$$

We then define the error transition as the stochastic matrix $E : S \to \Delta(S)$ sending each automaton of a type to the automaton of the same time, but choosing the state with 1/2 probability. [21] We denote $q^i(\omega)$ the state in S for player i to which player i transits given the current pair ω. Finally, we let $Q_\epsilon \in \mathcal{Q}$ to be:

$$\begin{aligned} Q_\epsilon((r^1, r^2), (s^1, s^2)) &= (1-\epsilon)^2 \text{ if } \forall i : r^i = q^i(s^1, s^2) \\ &= \epsilon(1-\epsilon) \text{ if for exactly one } i : r^i = q^i(s^1, s^2) \\ &= \epsilon^2 \text{ if } \forall i : r^i \neq q^i(s^1, s^2) \end{aligned}$$

G.1.2 Payoffs

If we let the payoff in the stage game at action profile $a \equiv (a^1, a^2)$ by $R(a)$ then we can define a payoff function $u : \Omega \to \mathbb{R}$ as

$$u(\omega) = u(s^1, s^2) = R(a^1(s^1)), a^2(s^2) \tag{A-17}$$

G.1.3 Value Function

Players choose an element in the set M, but the value function is defined for all elements in the set of pairs of extended states, Ω. The payoff to a player for each such $\omega \in \Omega$ is given by $V : \Omega \to \mathbb{R}$.

Lemma G.1. *The function $(\epsilon, \delta) \longrightarrow V(\cdot; \epsilon, \delta)$ is analytic, hence continuous and differentiable.*

Proof. The function V is the unique solution of the functional equation:

$$V = (1-\delta)u + \delta Q_\epsilon V \tag{A-18}$$

We may write $V(\omega; \delta, \epsilon, u)$ when we want to emphasize the dependence of V on these parameters. The inverse matrix $(I - \delta Q_\epsilon)^{-1}$ exists and therefore:

$$\begin{aligned} V(\cdot; \delta, \epsilon, u) &= (I - \delta Q_\epsilon)^{-1}(1-\delta)u \\ &= \sum_{k=0}^{+\infty} (\delta Q_\epsilon)^k (1-\delta) u \end{aligned}$$

---

1. $\forall a \in A : P(A; A, a) = 1$
2. $P(G_c; G_c, (C^1, C^2)) = 1, \forall_{a \neq (C^1, C^2)} P(G_d; G_c, (C^1, C^2)) = 1$
3. $\forall a \in A : P(G_d; G_d, a) = 1$
4. $\forall a^1 : P(T_c; T_c, (a^1, C^2)) = 1; P(T_d; T_c, (a^1, D^2)) = 1$
5. $\forall a^1 : P(T_c; T_d, (a^1, C^2)) = 1; P(T_d; T_d, (a^1, D^2)) = 1$

[21] More precisely,

1. $\forall a, \forall X \in \{G, T\}, \forall i \in \{c, d\} : E(X_c; X_i, a) = E(X_d; X_i, a) = 1/2$
2. $\forall a : E(A; A, a) = 1$.

The derivative of V with respect to the error parameter is

$$\begin{aligned}\frac{dV}{d\epsilon} &= -(I-\delta Q_\epsilon)^{-1}\delta\frac{dQ_\epsilon}{d\epsilon}(I-\delta Q_\epsilon)^{-1}(1-\delta)u \\ &= -(I-\delta Q_\epsilon)^{-1}\delta\frac{dQ_\epsilon}{d\epsilon}V\end{aligned}$$

□

The analysis of the function V is considerably simplified if we observe that Ω is partitioned into invariant sets under the transition Q_ϵ, as we do in the next section.

G.2 Invariant sets

It is clear that subsets of the set Ω are invariant under the transition P_ϵ. For example the set $\Omega_{AG} \equiv \{(A,G_c),(A,G_d)\}$ is invariant. The other eight sets are similarly naturally denoted. Overall we have a partition of Ω into

$$\mathcal{P}(\Omega) \equiv \{\Omega_{AG},\Omega_{AA},\Omega_{AT},\Omega_{GA},\Omega_{GG},\Omega_{GT},\Omega_{TA},\Omega_{TG},\Omega_{TT}\}. \quad \text{(A-19)}$$

each of which is invariant. Note that the cardinality of every element in this partition is either 2 or 4.

Correspondingly, the vector V is partitioned into component $V_i : i \in \mathcal{P}(\Omega)$, and each satisfies the equation (A-18) with (u, Q_ϵ) replaced by $(u_i, Q_{\epsilon,i})$; these equations can be solved and analyzed independently.

Lemma G.2. *The value function equation can be decomposed into nine independent equations, one for each of the invariant sets of the set* Ω*.*

1. $V(AA) = d$
2. $V(AG) = V(AT) = t(1-\delta(1-\epsilon)) + d\delta(1-\epsilon)$
3. $V(GA) = V(AT) = s(1-\delta(1-\epsilon)) + d\delta(1-\epsilon)$

Proof. The value of $V(AA)$ follows from the fact that the singleton $\{AA\}$ is invariant. The values of $V(AG)$ and $V(AT)$ follow as in the proof of lemma (G.3) below. □

The next lemma tells us that no matter what the probability of error, the profile (A,A) in the strategy choice game is an equilibrium:

Lemma G.3. *For all* $\epsilon > 0$, (A,A) *is a strict Nash equilibrium of the strategy choice game, hence a locally stable equilibrium of the PI and BR dynamics.*

Proof. The transition matrix restricted to the set Ω_{AG} is

$$\begin{array}{ccc} & G_cA & G_dA \\ G_cA & \epsilon & 1-\epsilon \\ G_dA & \epsilon & 1-\epsilon \end{array}$$

Using equation (A-18), we can solve for $V^1(G_cA)$ and find:

$$\begin{aligned} V^1(G_cA) &= (1-\delta(1-\epsilon))u(G_cA) + \delta(1-\epsilon)u(G_dA) \\ &= (1-\delta(1-\epsilon))s + \delta(1-\epsilon)d \end{aligned}$$

Therefore for all $\epsilon > 0$,

$$V^1(G_cA) < d = V^1(AA). \tag{A-20}$$

Since the transition matrix restricted to Ω_{AT} is the same as in (**??**),

$$V^1(T_cA) < V^1(AA). \tag{A-21}$$

□

G.3 Uniform Error

The simplification introduced in section (G.2) allows us to compute the payoff on the case of uniform error, that is when $\epsilon = 1/2$. This case gives us a boundary condition for the study of the dynamic behavior: in particular for example it will tell us when it is impossible that (G, G) is an equilibrium profile in the strategy choice game.

Lemma G.4. *When $\epsilon = 1/2$ the payoff in the strategy choice game (with $M \times M$ action set) is:*

	A	G	T
A	d, d	m_t, m_s	m_t, m_s
G	m_s, m_t	m_c, m_c	m_c, m_c
T	m_s, m_t	m_c, m_c	m_c, m_c

where

$$m_s \equiv (1-\delta)s + \delta\frac{s+d}{2}, m_t \equiv (1-\delta)t + \delta\frac{t+d}{2}, \tag{A-22}$$

$$m_c \equiv (1-\delta)c + \delta\left(\frac{c+s+t+d}{4}\right).$$

Thus at $\epsilon = 1/2$ the game has a unique Nash equilibrium, (A, A).

Proof. For any element $i \in \mathcal{P}(\Omega)$ defined in (A-19) we have that the transition restricted to the point in i is:

$$Q_{(1/2),i} = \frac{1}{\#i}U_i \tag{A-23}$$

where U_i is a square matrix of 1's of dimension $\#i$ (this being the cardinality of i). We denote $M_i x \equiv \frac{1}{\#i}\sum_{k=1}^{\#i} x_k$. The value function equation restricted to states in i is:

$$V_i = (1-\delta)u_i + \delta M_i V_i \tag{A-24}$$

The equality (A-24) implies

$$M_i V_i = M_i u_i \tag{A-25}$$

Now (A-24) and (A-25) give the formula for the value in terms of the payoffs:

$$V_i = (1-\delta)u_i + \delta M_i u_i \tag{A-26}$$

The rest follows from simple algebra.

For the last statement, from our assumption (A-3) on the stage game payoffs we conclude that $m_t > m_c$. We already know from lemma (G.3) that $d > m_s$ holds for any δ. Thus A is a dominant action. □

G.4 Small Errors

We know that for "large" $\epsilon = 1/2$ the strategy choice game has a unique equilibrium, defection in every round irrespective of history for both players $((A, A))$. The dynamic behavior for small ϵ on the sides $\mu(T) = 0$ and $\mu(G) = 0$ is similar to that of the no error model, with a unique steady state the changes continuously in ϵ around $\epsilon = 0$. Instead, in the portion of the interior of the simplex where A is not the attractor, and on the line $\mu(A) = 0$ the behavior is radically different, because the only stable states are either T or G. Consequently in this case we have well defined basins of attraction for the two strategies G and T and we can compare the two basins for the lenient strategy T and the strict strategy G.

"Small ϵ" means in our analysis smaller than a critical value (which may not be numerically small), that we introduce now. We let:

$$\bar{\epsilon} \equiv \sup\{\epsilon : V^1(GG) > \max\{V^1(TG)), V^1(AG)\} \tag{A-27}$$
$$\& V^1(TT) > \max\{V^1(AT)), V^1(GT)\}\}$$

Lemma G.5. *For any payoff of the stage game, $\bar{\epsilon} > 0$. If it is finite, then the value is achieved. There are payoffs satisfying the standing assumption on stage game payoffs (A-3) for which $\bar{\epsilon}$ is finite.*

Proof. Note that at $\epsilon = 0$,

$$\begin{aligned} V^1(GG) = & \quad V^1(TG) \\ = & \quad c \\ > & \quad V^1(AG) \\ = & \quad t(1-\delta) + d\delta \end{aligned}$$

where the strict inequality follows from our assumption (A-7). Thus we conclude that $\bar{\epsilon} > 0$ from lemma (G.1). The second claim also follows from lemma (G.1). From last claim follows from lemma (G.4), which gives as simple sufficient condition for $\bar{\epsilon}$ to be finite. □

In particular $\bar{\epsilon}$ is finite for the stage game in our experimental design.

Note that when $\epsilon < \bar{\epsilon}$ then the best response to strategies supported in $\{G, T\}$ for player 1 is supported in Nash equilibria in the game with strategy profiles $M \times M$ which give

Lemma G.6. *We consider either the proportional imitation or the best response model. For $0 < \epsilon < \bar{\epsilon}$:*

1. *the sides of the simplex are invariant;*
2. *on the sides $\mu(G) = 0$ (and, respectively, $\mu(T) = 0$) there is a unique steady state corresponding to the mixed strategy equilibrium of the game when strategies are restricted to $\{A, T\}$ for both players (and respectively to $\{A, G\}$);*
3. *there is a unique steady state on the side $\mu(A) = 0$, and this is determined in the limit $\epsilon \to 0$ by the ratio:*

$$\mu(G)^* = \frac{V^1_\epsilon(TT) - V^1_\epsilon(GT)}{V^1_\epsilon(TT) - V^1_\epsilon(GT) + V^1_\epsilon(GG) - V^1_\epsilon(TG)} \tag{A-28}$$

where V^1_ϵ is the derivative with respect to ϵ.

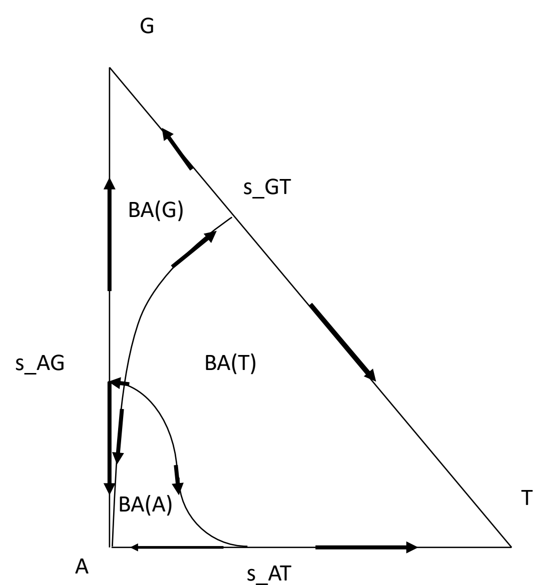
Proof. For point (3), note first that for $\epsilon > 0$ then the steady state is determined by a frequency of the strategy G, $\mu(G)^*$ such that

$$\mu(G)^* = \frac{V^1(TT) - V^1(GT)}{V^1(TT) - V^1(GT) + V^1(GG) - V^1(TG)} \tag{A-29}$$

As $\epsilon \to 0$ both numerator and denominator in the right hand side of (A-29) tend to zero. Now the conclusion follows from l'Hôpital rule. □

Figure (A.6) below illustrates the dynamic behavior of the proportional imitation model when small errors occur. There are seven steady states. The three corresponding to the pure strategy profiles are locally stable. The basin of attractions, indicated as $BA(G)$ (Basin of attraction of G) and so on), are delimited by the curved lines describing the manifolds departing from the unique steady state in the interior of the triangular region and the sides. The other four steady states are unstable.

Figure A.6: **Proportional Imitation dynamics with error.** "Small" error ($\epsilon < \bar{\epsilon}$). There are three stable states corresponding to the pure strategy profiles (A, A), (G, G) and (T, T). The steady state in the interior of the simplex is unstable, and a Nash equilibrium. The remaining three are mixed Nash and saddle point unstable.



H Errors and frequency of strategies

We now turn to the main question we posed, namely the relationship between the probability of error (modeled by the ϵ parameter) and the frequency of strict and lenient strategies. We consider the natural strictness order:

$$A \succ G \succ T \tag{A-30}$$

We denote the steady states (when they exist) on the sides of the simplex $\{A, T\}$, $\{A, G\}$ and $\{G, T\}$, as s_{AT}, s_{AG} and s_{GT} respectively. For illustration we refer to figures (A.5) and (A.6). These are real numbers in the unit interval, equal to the size of the basin of attraction of the stricter strategy in the subset. When we want to emphasize the dependence of the parameters, we write $s_{ij}(\epsilon, \delta, u)$.

H.1 Gains and Basins of Attraction

Some very elementary but useful concepts are needed here. Consider a symmetric game with action set $\{A, B\}$ where each action is a best response to itself, that is, the *gains* $G(A) \equiv u(A, A) - u(B, A)$ and $G(B) \equiv u(B, B) - u(A, B)$ satisfy $G(A) > 0$ and $G(B) > 0$. THis game has two pure strategy Nash equilibria and a mixed strategy one; the mixed strategy equilibrium has

$$\mu(A) = \frac{G(B)}{G(A) + G(B)}. \tag{A-31}$$

In this simple two-actions game the basin of attraction of A is the set $\{(p, 1-p) : 1 \geq p > \mu(A)\}$, so the *size* of the basin of attraction of A,

$$SBA(A) = \frac{G(A)}{G(A) + G(B)}. \tag{A-32}$$

that is, as intuitive, the size of the basin of attraction of a strategy is proportional to the relative gain of one action over the gain of the other.

I Proportional Imitation Dynamics

An increase in the values of s_{AT}, s_{AG} corresponds to an increase in the basin of attraction of A at the expense of those of G and T; an increase in the value of s_{GT} leaves the relative share of A unchanged, but it increases that of G at the expense of T.

The following proposition summarizes what we know for the two extreme value of $\epsilon = 0$ and $= 1/2$:

Proposition I.1. *Under proportional imitation,*

$$s_{AG}(0, \delta, u) = s_{AT}(0, \delta, u) = \frac{(1-\delta)d}{c - (1-\delta)t - \delta d + (1-\delta)d} \tag{A-33}$$

$$s_{AG}(1/2, \delta, u) = s_{AT}(1/2, \delta, u) = s_{GT}(1/2, \delta, u) = 1 \tag{A-34}$$

So at $\epsilon = 0$ the basin of A is the triangular region below the straight-line segment joining s_{AG} and s_{AT}. As ϵ reaches the value $1/2$ the entire simplex (except the side between G and T) converges to A. The strategy A may become the only strategy surviving for values of the error probability smaller than $1/2$, as we are going to see immediately.

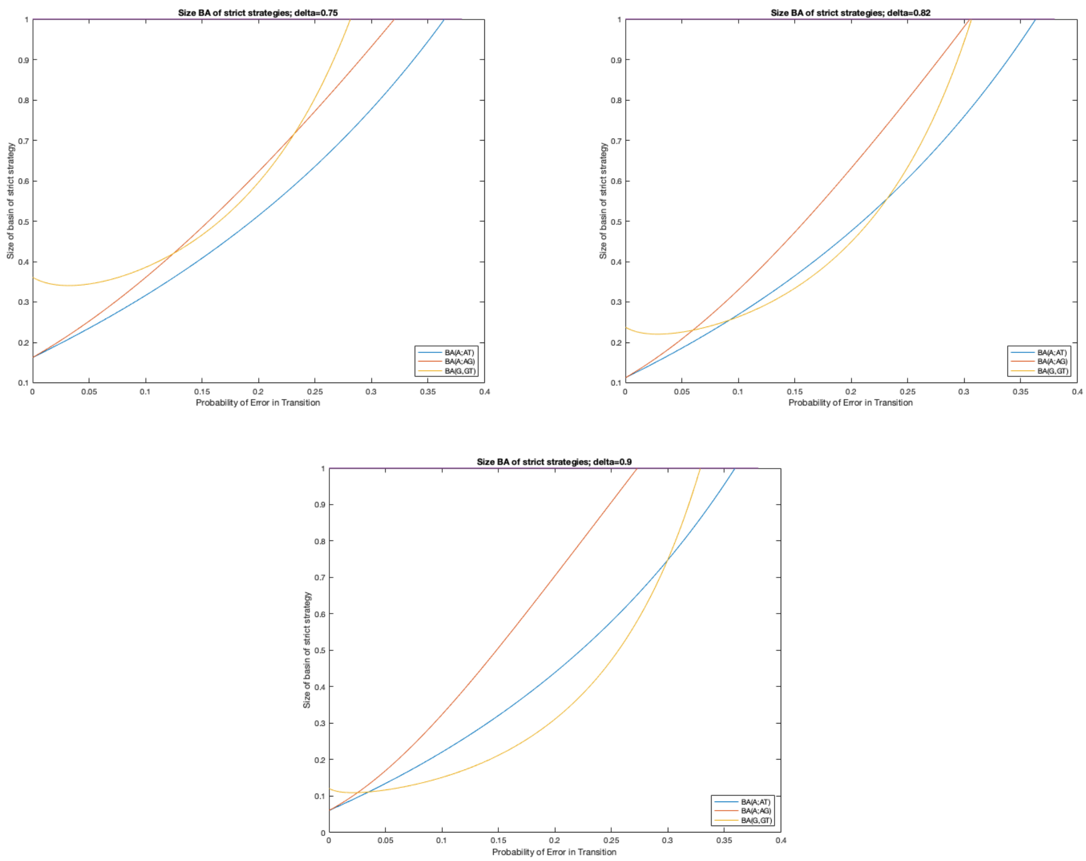
I.1 Payoff of experiment

To see how the size of the basin of attraction changes when the stage game payoffs are those we used in the experiment ($c = 48$, $s = 0$, $t = 50$, $d = 25$), we can use the lemmas (G.1) and (G.2) to compute the value function and analyze equilibria and dynamic behavior.

We refer to Figure (A.7), which describes how the change for different values of ϵ (moving along the x-axis in each panel) and δ.

The first conclusion is:

Figure A.7: **Size of basin of attraction of strict strategies.** Top to bottom panel: values of δ equal to 0.75, 0.82, 0.9 respectively.



Conclusion I.2. *With sufficiently high continuation probability and small error, the basin of attraction of defection (A) is arbitrarily small.*

The values s_{AG} and s_{AT} are monotonically increasing in ϵ, the values s_{GT} is increasing for most of the range.

Note that for all values of δ the last value to reach the upper range of one is s_{AT}. The intermediate value of $\delta = 0.82$ marks the boundary of an interesting division of the set of δ's.

Conclusion I.3. *For any value of the continuation probability, as the error becomes larger, the basin of attraction of lenient strategies (G and T) vanishes. Within the more lenient strategies, the relative weight of T declines compared to that of G as the error probability increases.*

For lower values ($\delta < \hat{\delta}$) the basin of attraction of T disappears entirely as s_{GT} collapses into the vertex T. That is, with interactions repeating with lower continuation probability (lower δ's) but high probability of errors the strategy T does not survive. For higher values ($\delta > 0.82$) both G and T survive, but lose frequency at the expense of A. In conclusion,

Conclusion I.4. *The lenient strategy T can survive only with low errors or with high probability of continuation.*

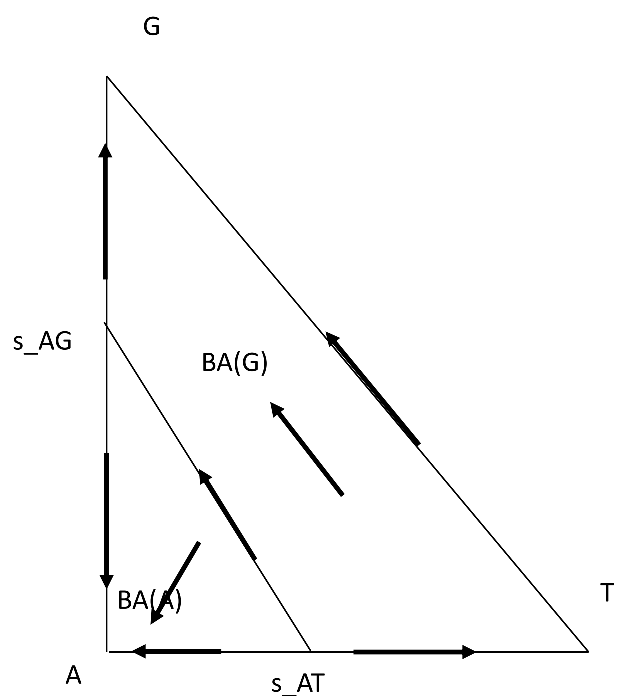
Figures (A.8) and (A.9) illustrate the dynamics of the proportional imitation for values above and below the threshold, and payoff equal to those used in the experiment. By "first transition" we refer here to the first disappearance of a steady state on the sides of the simplex. We denote $\hat{\delta}$ the value of δ at which s_{GT} and s_{AT} disappear at the same value of ϵ; with the payoffs used in the experimental design, $\hat{\delta} = 0.82$.

Conclusion I.5. *There is a $\hat{\delta}$ such that*

1. *if $\delta > \hat{\delta}$ then as ϵ increases the interior of the simplex is split first into three regions (with attractors A, G and T respectively); then two (with attractors A and T) and finally only one region) (with attractors A)*

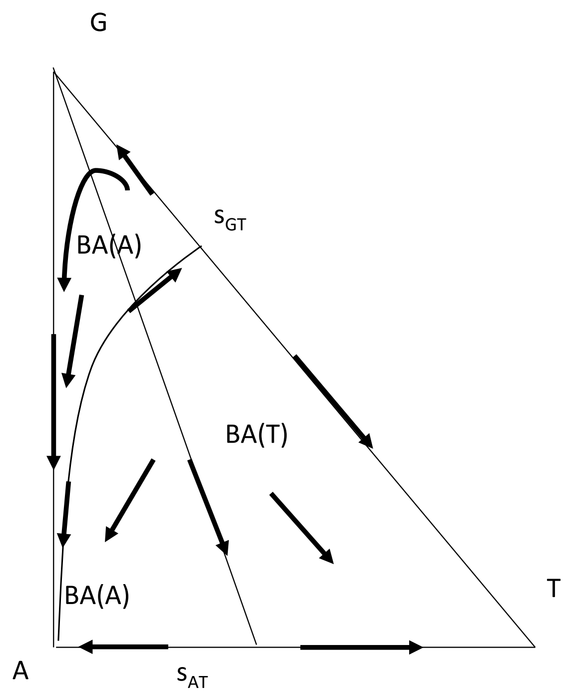
2. *if $\delta < \hat{\delta}$ basins of attraction are the same for small and high ϵ; but in the intermediate region the attractors are A and G*

Figure A.8: **Flow after the first transition: Low** δ**.** Here $\epsilon > \bar{\epsilon}$), but the two values are close. Also $\delta < \hat{\delta}$, hence the steady state s_{GT} disappears first. There are only two sub-regions of the interior of the projection of the simplex, with attractors A and G.



In Figure (A.8), since all the steady states are isolated, an appropriately modified Poincare-Hopf index can be calculated. The index of the three stable pure strategies are all 1; the others (the two mixed strategies on the sides $\{A, T\}$ and $\{A, G\}$), with an overall index 1. Note that the pure strategy profile (T, T) is a steady state of the proportional imitation, and a Nash equilibrium of the game restricted to $\{A, T\}$ but it is not a Nash equilibrium of the complete game.

Figure A.9: **Flow after the first transition: high δ.** Here $\epsilon > \bar{\epsilon}$), but the two values are close. In this case $\delta > \hat{\delta}$, hence the steady state s_{AG} disappears first. There are only two sub-regions of the interior of the projection of the simplex, with attractors A and T.



In Figure (A.9) also all the steady states are isolated, and thus an appropriately modified Poincare-Hopf index can be calculated. The index of the two stable pure strategies (A and T) are 1; the mixed strategy in the interior of the simplex has index 1. The index of the two steady states on the sides (s_{AT} and s_{GT}) is -1. The index of G is 0.

J Best Response Dynamics

We now consider the dynamic evolution when the evolution follows the best response dynamics. To understand the difference between this and the proportional imitation dynamics, it is useful to keep in mind that there may be steady states on the boundary of the simplex for the PID that are not Nash equilibria of the entire game. To be precise, let $NE(\{A, G, T\})$ be the set of Nash equilibria of the strategy choice game and, for any two-strategy subset $\{r, s\}$ of the set $\{A, G, T\}$, and let $N(\{r, s\})$ the Nash equilibria of the reduced game where players can only choose from $\{r, s\}$. It may occur that a steady state at the boundary for the PID is not a Nash, although it may be a Nash equilibrium of the game restricted to the strategies that have positive probability at that steady state.

For example, in the game induced by $\epsilon = 0.35$ ($\delta = 0.9$) in Table (A.20) the strategy T is weakly dominated by G and $V(GT) > V(TT)$ so the steady state s_{AT} of the PID is not Nash. The same is true for pure strategies: for example, the strategy G is a steady state in the game induced by

$\epsilon = 0.3$ ($\delta = 0.9$) in Table (A.20), the pure strategy G is a steady state of the PID (this is illustrated in Figure A.9), but is not a Nash equilibrium of the complete game, because $V(AG) > V(GG)$.

Proposition J.1. *With best response dynamics,*

1. *for $\epsilon < \overline{\epsilon}$, all three strategies have a basin of attraction; the intersection of the boundaries of the regions with the sides of the simplex are the same as in the proportional imitation dynamics*

2. *The conclusions (I.2), (I.3), (I.4) and (I.5) hold in the case of BRD as well; inparticular, for $\epsilon > \overline{\epsilon}$ only two strategies survive in the long run (that is, they have a basin of attraction): they are $\{A, G\}$ for low delta ($\delta < \hat{\delta}$), and $\{A, T\}$ for high δ ($\delta > \hat{\delta}$).*

Note that the lines marking the boundary of the basins of attraction in the BRD and PID may be different (those in BRD are straight lines) but the end points are common, hence qualitatively the basins are the same.

J.1 Best Response Dynamics

The best response dynamic allows us to identify clearly the basin of attraction of the three strategies by simple inspection of the phase portrait. This is done in Figure (10) that we have already discussed in our introductory discussion, and we refer the reader to that section.

A Payoff Tables with Transition Errors

For completeness we report the value of the matrix V for the values of δ relative to Figure (A.7), top and bottom panel, for various values of ϵ. The entries illustrate the transition along different types of equilibria and attractors.

In Table (A.19), with $\delta = 0.75$, as ϵ crosses the first threshold between 0.25 and 0.3, action T becomes dominated by G, so s_{GT} disappears. Figure (A.8) illustrates the dynamics in this situation. After the second threshold, $V(A, G) > V(G, G)$ and thus s_{AG} disappears. After the last, A becomes dominant.

Table A.19: V **matrix,** $\delta = 0.75$**,** $\epsilon = 0.25, 0.30, 0.35, 0.40$**.**

	A	G	T
A	25.0000	35.9375	35.9375
G	19.3125	37.6125	39.1168
T	19.3125	37.2846	39.1962
	A	G	T
A	25.0000	36.8750	36.8750
G	18.8250	37.3221	38.6771
T	18.8250	37.0298	38.6384
	A	G	T
A	25.0000	37.8125	37.8125
G	18.3375	37.1935	38.3034
T	18.3375	36.9510	38.1969
	A	G	T
A	25.0000	38.7500	38.7500
G	17.8500	37.1718	37.9620
T	17.8500	36.9941	37.8411

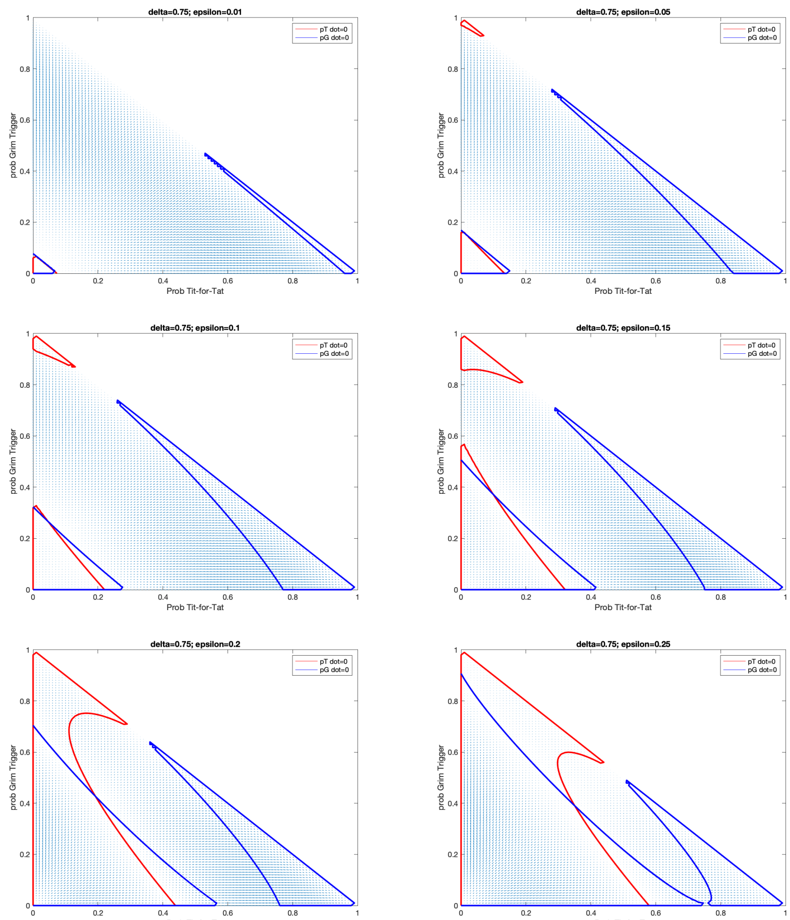
In Table (A.20), after the first threshold, $V(A,G) > V(G,G)$ and thus s_{AG} disappears first. After the second threshold, G dominates T and thus s_{GT} disappears at this point. Figure (A.9) illustrates this case.

Table A.20: V **matrix,** $\delta = 0.9$**,** $\epsilon = 0.25, 0.30, 0.35, 0.40$**.**

	A	G	T
A	25.0000	33.1250	33.1250
G	20.7750	33.5591	35.7744
T	20.7750	33.1806	36.1957
	A	G	T
A	25.0000	34.2500	34.2500
G	20.1900	33.7727	35.7489
T	20.1900	33.4154	35.8681
	A	G	T
A	25.0000	35.3750	35.3750
G	19.6050	34.0813	35.6889
T	19.6050	33.7695	35.6246
	A	G	T
A	25.0000	36.5000	36.5000
G	19.0200	34.4359	35.5754
T	19.0200	34.1972	35.4377

B Vector Field with Replicator Dynamics

Figure A.10: **Basin of attraction of A, G and T, with transition error and Proportional Imitation dynamics.** The probability of error in transition is as displayed at the top of each panel, and is ranging from 1 per cent to 25 per cent. Payoff and discount factor ($\delta = 0.75$) are as in our experimental design.



C Errors in Action and Transition

In this section we analyze the more complete (and more complex) model in which errors of both types (in action choice and in transition) are possible.

The main innovation with respect to the analysis in the main text is the introduction of the error in action transition matrix. We denote (just as ϵ was the probability of an independent error in the transition to the new state of the automaton) by η the probability of an error in the choice of the action at a state in the automaton. The set of action profiles is as usual $A \equiv A^1 \times A^2$. Let $Pr(a; \omega, \eta)$ denote the probability of choice of the action profile a by the two players when the current state is ω and the probability of an error in action choice is η. The action choice with errors at a state is a stochastic matrix $A_\eta : \Omega \to \Delta(\Omega \times A)$ defined by

$$A_\eta(\omega)(\omega', a) \equiv \delta_\omega(\omega')Pr(a; \omega, \eta) \tag{A-35}$$

Note that the ω coordinate in the image space of A is only required as a placeholder. This role turns out to be essential in the next step, the definition of $Q_{\epsilon,\eta}$ in equation (A-36). To illustrate the definition of A_η consider for example:

$$A_\eta(G_c, G_d)(G_c, G_d, C^1, D^1) = (1 - \eta)^2$$

The transition with errors $T_\epsilon : \Omega \times A \to \Delta(\Omega)$ is defined by taking $T_\epsilon(\omega, a)(\omega')$ as the probability that the next period state is ω' given that the current state is ω and current action profile is a. Overall the stochastic matrix $Q_{\epsilon,\eta} \in \mathcal{S}(\Omega, \Omega)$ is the composition of the two transitions:

$$Q_{\epsilon,\eta}(\omega'; \omega) \equiv \sum_{a \in A} A_\eta(\omega)(\omega, a)T_\epsilon(\omega, a)(\omega') \tag{A-36}$$

We denote by $u_\eta(a)$ the one period payoff when the intended action profile is a but errors in action choice are possible and occur independently for the two players with probability η. To illustrate, if the intended action profile is (C^1, D^2) then $u_\eta(C^1, D^2) = (1-\eta)^2 s + \eta(1-\eta)(d+c) + \eta^2 t$. We also let $V_{\epsilon,\eta}(\omega)$ the value function at the state ω

$$V_{\epsilon,\eta} = (1 - \delta)u_\eta + \delta Q_{\epsilon,\eta} V_{\epsilon,\eta} \tag{A-37}$$

C.1 The Nash Equilibrium Set

The analysis of the properties of the value function presented in the main text holds with little adjustments in the current case where errors in actions and transition are possible. The following lemma holds:

Lemma C.1. *The function $(\epsilon, \eta, \delta) \longrightarrow V(\cdot; \epsilon, \eta, \delta)$ is analytic, hence continuous and differentiable.*

The decomposition of the state space described in the main text, section (G.2) holds in the current case as well. Similarly, with easy computations, one gets:

Lemma C.2. *The value function equation can be decomposed into nine independent equations, one for each of the invariant sets of the set Ω.*

1. $V_{\epsilon,\eta}(AA) = u_\eta(DD)$
2. $V_{\epsilon,\eta}(GA) = (1 - \delta(1 - \epsilon))u_\eta u(CD) + \delta(1 - \epsilon)u_\eta(DD)$

3. $V_{\epsilon,\eta}(AG) = (1-\delta(1-\epsilon))u_\eta u(DC) + \delta(1-\epsilon)u_\eta(DD)$

Note that

$$\begin{aligned} u_\eta(DD) &= (1-\eta)^2 d + \eta(1-\eta)(s+t) + \eta^2 c \\ u_\eta(CD) &= (1-\eta)^2 s + \eta(1-\eta)(d+c) + \eta^2 t \\ u_\eta(DC) &= (1-\eta)^2 t + \eta(1-\eta)(d+c) + \eta^2 s \end{aligned}$$

In the case of the two errors the best response to A is A for the interesting values of the parameter η:

Lemma C.3. *For all $\epsilon > 0$ and $\eta < 1/2$, (A, A) is a strict Nash equilibrium of the strategy choice game, hence a locally stable equilibrium of the PI and BR dynamics.*

Proof. An easy computation gives:

$$V_{\epsilon,\eta}(AA) - V_{\epsilon,\eta}(GA) = (1-2\eta)(1-\eta)(d+t-s-c).$$

□

Hence in the case of the two errors too A survives for all values of the parameters.

C.2 Basin of Attraction and Error Rates

From our analysis it is clear that the behavior of the basin of attraction as function of the two error rates will broadly follow a behavior similar to that already observed in the case of the simple error in transition, examined in the main text. Figures (A.11) and (A.12) report the sizes of the basin as function of the two error rates. These are the three dimensional versions of the figure (A.7).

One can consider also the exact analogue of figure (A.7), but for the error in action choice, by setting the transition probability to zero. This makes clear that the basin of attractions of the strict strategies in the sides of the simplex (analyzing the size of the basin when players are playing only $\{G, T\}$, $\{A, T\}$ and $\{A, T\}$) is strictly increasing in the probability of the error in action choice.

An important difference between the effect of error in transition and error in action is the following. Even with no error in action a large enough error probability in transition will reduce the strategy surviving to A: this is clear considering the projection on the zero error in action plane. Instead, with no error in transition only the

In the case of the subset $\{A, T\}$, at zero transition error the effect of the error in action is still strictly increasing, but of limited size: at $\delta = 0.75$ the largest size of the basin of attraction of A is approximately 25 per cent; at $\delta = 0.9$ the maximum is smaller than 10 per cent.

Figure A.11: **Size of basin of attraction of strict strategies with two types of error.** Probability of error in transition and action choice as displayed. $\delta = 0.75$. Top to bottom panel: values of the basin of attraction in the indicated subsets of strategies (basin of G in GT, A in AG, A in AT, respectively).

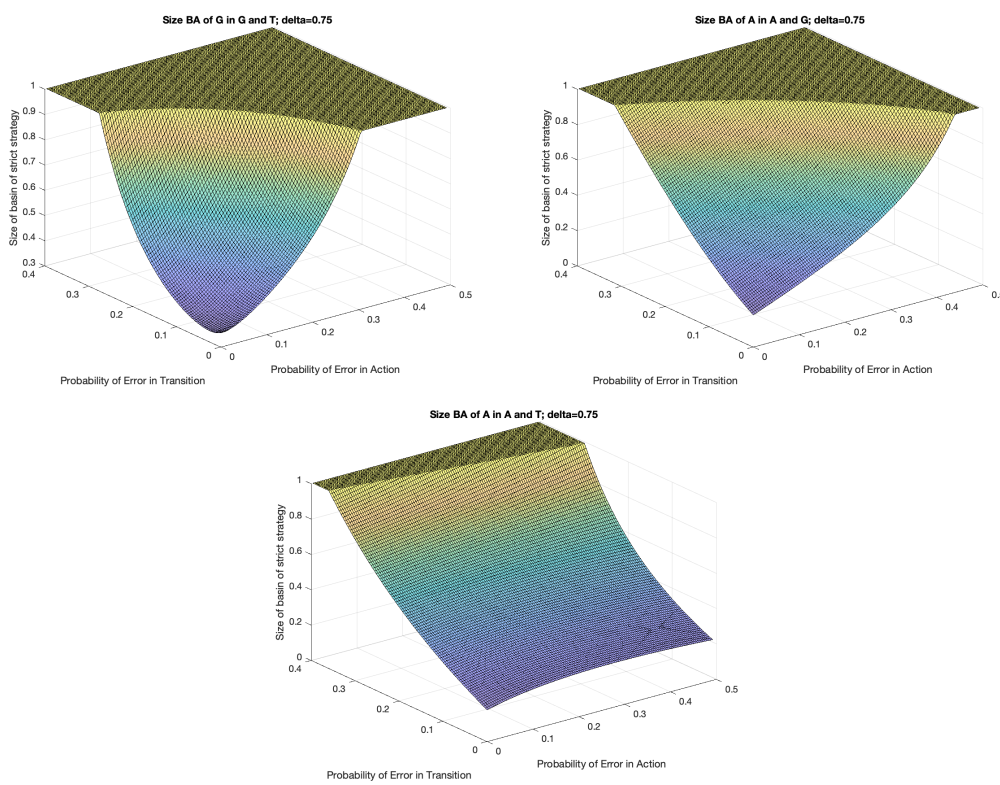
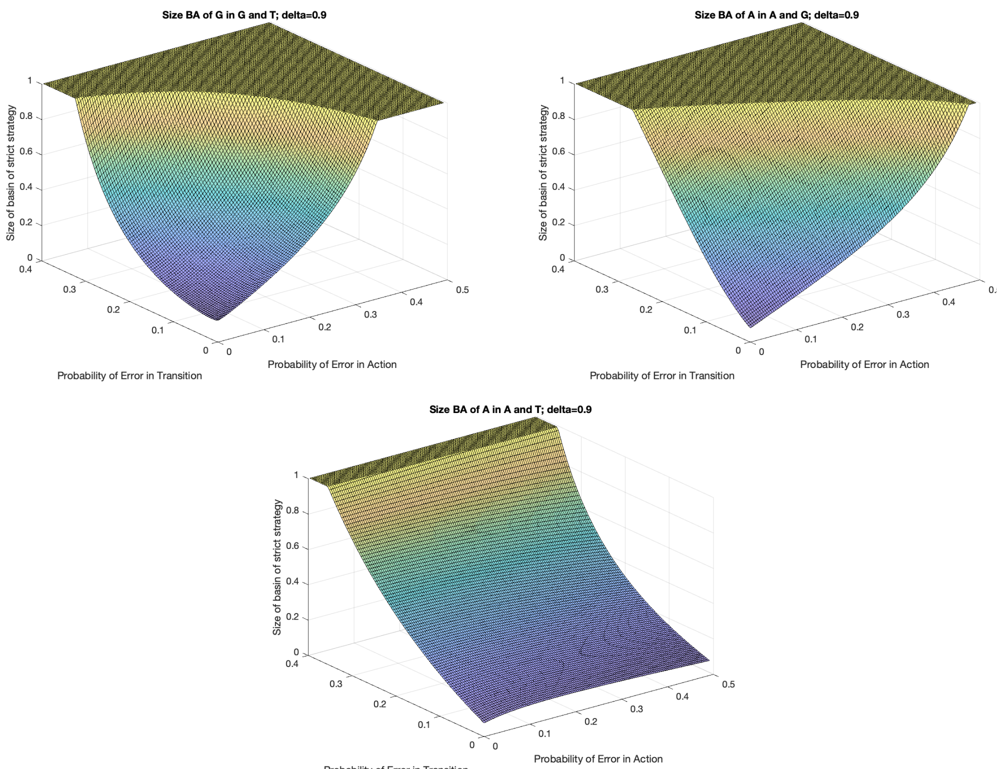


Figure A.12: **Size of basin of attraction of strict strategies with two types of error.** Probability of error in transition and action choice as displayed; $\delta = 0.9$. Top to bottom panel: values of the basin of attraction in the indicated subsets of strategies (basin of G in GT, A in AG, A in AT, respectively).



D Learning Model

We now consider an explicit model of the learning process that will provide a link between the model we have discussed so far and the analysis of the data performed in the main text, and in particular with the method presented in section (4).

D.1 A Population Model

We consider a model in which a population of players; each player is indexed by an index $i \in [0, 1]$, and has a point in time a strategy profile he adopts, call it $s^i(t)$. Such assignement of strategy to each player induces a probability distribution on strategies, where we call as in the main text $\mu(s,t)$ the frequency of the strategy s) (that is, the fraction of the i players who have adopted s at t). Players know the distribution μ and can compute the best response, but for some reason they cannot adopt the best response when they like.

D.1.1 Best Response

We apply this setup to provide a justification of the Best Response dynamic (see for example Sandholm (2007)), that we described in the main text as:

$$\forall s \in S : \frac{d\mu(s,t)}{dt} \in \lambda \left(BR(\mu(\cdot,t))(s) - \mu(s,t) \right). \tag{A-38}$$

we have added a minor modification, a $\lambda \in \mathbb{R}_+$; the reason will become clear in a moment. In the time interval dt (dt small) a fraction λdt of the population can revise their strategy; the complement $1 - \lambda dt$ cannot. Those who can, adopt the best response to the current $\mu(\cdot, t)$; each player does so taking the current μ as given, and ignoring (correctly, since he is negligible) his effect on the frequency of strategies. If we denote by BR the set of optimal mixed strategies, for every m we have

$$\mu(s, t + dt) \in (1 - \lambda dt)\mu(s, t) + \lambda dt BR(\mu(\cdot, t))(s) \tag{A-39}$$

which in the limit gives the differential inclusion (A-38).

We now apply the population model to the learning model used in the main text.

D.2 The learning model

Each player has a belief on the distribution of strategies in the population. The belief has the same form for all players, and is a Dirichelet distribution of the three dimensional simplex, $\Delta(\{A, G, T\})$; player i has a concentration parameter $\alpha^i \in \mathbb{N}^3$; so the density of the belief of that player is:

$$D(\mu, \alpha^i) = \frac{1}{B(\alpha^i)} \mu(A)^{\alpha^i_A - 1} \mu(G)^{\alpha^i_G - 1} \mu(T)^{\alpha^i_T - 1} \tag{A-40}$$

where B is the beta function.

The assignment of the belief described by α^i to each player induces a population distribution over belief of players on the strategy of others, described by a probability distribution on the countable set $\mathbb{N}^3$; with generic term π.

D.2.1 Matching and Playing the Game

At time t, a fraction λdt of the population is randomly selected to play the game. This sub-population is representative of the total population, so the distribution on $\mathbb{N}^3$ in it is the same as in the total population. For convenience, we will consider in the following the extension of the measure π to a measure on $\mathbb{Z}^3$, set equal to 0 on all three dimensional vectors of integers that have a negative value in some coordinate; we keep the same symbol:

$$\pi \in \Delta(\mathbb{Z}^3) \tag{A-41}$$

D.2.2 Players' best response

We now consider players in the selected sub-population. Given the distribution on the strategy of the other selected players, each player computes and chooses with no limitation a mixed strategy in the set of his best responses, given his belief indexed by α:

$$BR(\alpha) = argmax_{\sigma \in \Delta(S)} E_{D(\cdot,\alpha)} u(\sigma, \cdot). \tag{A-42}$$

By the property of the Dirichelet distribution with α that the mean of the $\mu_s, s \in S$ is

$$\begin{aligned} E_{D(\cdot,\alpha)}\mu(s) &= \frac{\alpha_s}{\sum_{r \in S} \alpha_r}. \\ &\equiv M(s;\alpha) \end{aligned} \tag{A-43}$$

When convenient write $M(\cdot, \alpha)$ simply as $M(\alpha)$. The best response set of a player with belief indexed by α is:

$$BR(\alpha) = argmax_{\sigma \in \Delta(S)} \sum_{s \in S} M(s;\alpha) u(\sigma, s). \tag{A-44}$$

When we add over $\mathbb{N}^3$ with weights given by π, the best response of each player we get an element $\phi(\cdot;\pi) \in \Delta(S)$ which is the true distribution in the population of the strategies. Note that the function ϕ depends on the value function of the repeated game at the corresponding error rate. We will make this dependence explicit later on when we need to study its effects, but we ignore it for the moment for clarity of notation. [22]

D.2.3 Discrete Time Evolution of Belief Distribution

We consider first the evolution in discrete time. In each period, all players are randomly matched with probability corresponding to the frequency.

Proposition D.1. *The fraction of each α belief in the model described in section (D.2) follows the equation:*

$$\pi(\alpha, k+1) = S\pi(\cdot, k), k = 0, 1, \ldots. \tag{A-46}$$

where S defined in (A-48).

[22] We assume that when the best response of a player with belief α is not a pure strategy, then players choose according to the uniform distribution over the best response set, so when we aggregate over the sub-population of such players we get the expected value of the strategy choice. In detail, we define:

$$\begin{aligned} \phi(s;\pi) \equiv \pi(\{\alpha : BR(\alpha) = \{s\}) + \sum_{r \neq s} \frac{1}{2} \pi(\{\alpha : BR(\alpha) = \{s, r\}) + \\ \frac{1}{3}\pi(\{\alpha : BR(\alpha) = \Delta(S)) \end{aligned} \tag{A-45}$$

Proof. Let 1_s denote the three dimensional vector equal to 1 at the sth coordinate, and 0 otherwise. Given a $\pi \in \Delta(\mathbb{N}^3)$, a players hold a belief α next period if and only if in the current period he holds a belief $\alpha - 1_s$ and meets an opponent playing s, which happens with probability $\phi(s, \pi)$. Each player then updates his belief; since priors are Dirichelet, he changes the α^i to the new value:

$$\alpha_s^{i,\prime} = \alpha_s^i + \delta_s(b^i) \tag{A-47}$$

where δ_s is the indicator function. [23] Define:

$$(S\pi)(\alpha) \equiv \sum_{s \in S} \pi(\alpha - 1_s)\phi(s, \pi) \tag{A-48}$$

Recall our discussion before (A-41), so the definition (A-48) is meaningful even when, in $\alpha_s = 0$. Equation (A-46) follows. □

D.2.4 Continuous Time Evolution of Belief Distribution

We now show that the time evolution of the fraction of beliefs in the population follows a dynamic very similar to the one described by the best response presented in section (D.1.1), with the distribution on beliefs π replacing the distribution on strategies μ:

Proposition D.2. *The time derivative of the fraction of each α belief in the model described in section (D.2) follows the equation:*

$$\frac{d\pi(\alpha, t)}{dt} = \lambda \left(\sum_{s \in S} \pi(\alpha - 1_s)\phi(s, \pi(\cdot, t)) - \pi(\alpha, t) \right) \tag{A-49}$$

Proof. The players who are matched to play observe a strategy b^i of the opponent with probability $\phi(b^i; \pi)$. Each player in the sub-population updates his belief according to (A-47). The new population after the time interval dt is a combination of the population of players that did not play, that is a fraction $1 - \lambda dt$, with frequency $\pi(\cdot, t)$ unchanged; and the sub-population of selected players, a fraction λdt, with frequency $S\pi(\cdot, t)$. Thus, the frequency next period is

$$\pi(\alpha, t + dt) = (1 - \lambda dt)\pi(\alpha, t + dt) + \lambda dt S\pi(\cdot, t),$$

and therefore (A-49) follows. □

The analysis of the evolution over time is more difficult to visualize than it is in the simple two dimensional case of the best response dynamic, but the logic is the same. In particular consider the best response function ϕ as depending on the value function V for a given vector of parameters, $(\epsilon, \eta, \delta, u)$. As the error in action becomes large, the best response assigns for the same $\pi(\cdot, t)$ a larger weight to the strategy A, until the frequency converges to the consensus on A.

[23] That is, $\delta_s(b^i) = 1$ if $s = b^i$ and $= 0$ otherwise.

References

Alaoui, L. and Penta, A. (2015). Endogenous depth of reasoning, *The Review of Economic Studies* **83**(4): 1297–1333.

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.

Bock, O., Baetge, I. and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool, *European Economic Review* **71**(C): 117–120.

Dal Bó, P. (2005). Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games, *American Economic Review* **95**(5): 1591–1604.

Dal Bó, P. and Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence, *American Economic Review* **101**(1): 411–429.

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.

Dreber, A., Rand, D. G., Fudenberg, D. and Nowak, M. A. (2008). Winners do not punish, *Nature* **452**: 348–351.

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

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* **10**(2): 171–178.

Fudenberg, D., Rand, D. G. and Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world, *American Economic Review* **102**(2): 720–49.

Gilboa, I. and Matsui, A. (1991). Social stability and equilibrium, *Econometrica* **59**(3): 859–67. **URL:** *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:59:y:1991:i:3:p:859-67*

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy* **124**(6): 1619–1676.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects, *American Economic Review* **92**(5): 1644–1655.

Hyndman, K., Ozbay, E. Y., Schotter, A. and Ehrblatt, W. Z. (2012). Convergence: an experimental study of teaching and learning in repeated games, *Journal of the European Economic Association* **10**(3): 573–604.

John, O. P., Donahue, E. M. and Kentle, R. L. (1991). *The big five inventory: versions 4a and 54*, Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P. and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy, *in* R. W. R. Oliver P. John and L. A. Pervin (eds), *Handbook of Personality: Theory and Research*, Guilford Press New York, NY, pp. 114–158.

Larson, G. E., Saccuzzo, D. P. and Brown, J. (1994). Motivation: Cause or confound in information processing/intelligence correlations?, *Acta Psychologica* **85**(1): 25–37.

Muthukrishna, M., Morgan, T. J. and Henrich, J. (2016). The when and who of social learning and conformist transmission, *Evolution and Human Behavior* **37**(1): 10–20.

Proto, E., Rustichini, A. and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions, *Journal of Political Economy* **0**(ja): null.

Sandholm, W. H. (2007). *Evolutionary Game Theory*, University of Wisconsin.

Schlag, K. (1998). Why imitate, and if so, how?,: A boundedly rational approach to multi-armed bandits, *Journal of Economic Theory* **78**(1): 130–156.
URL: *https://EconPapers.repec.org/RePEc:eee:jetheo:v:78:y:1998:i:1:p:130-156*

Vostroknutov, A., Polonio, L. and Coricelli, G. (2018). The role of intelligence in social learning, *Scientific reports* **8**(1): 6896.

Weibull, J. W. (1997). *Evolutionary Game Theory*, MIT Press Books, The MIT Press.