DISCUSSION PAPER SERIES

DP14349

INTELLIGENCE, ERRORS AND STRATEGIC CHOICES IN THE REPEATED PRISONERS' DILEMMA

Eugenio Proto, Aldo Rustichini and Andis Sofianos

LABOUR ECONOMICS

PUBLIC ECONOMICS

ISSN 0265-8003

INTELLIGENCE, ERRORS AND STRATEGIC CHOICES IN THE REPEATED PRISONERS' DILEMMA

Eugenio Proto, Aldo Rustichini and Andis Sofianos

Discussion Paper DP14349
Published 23 January 2020
Submitted 21 January 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

INTELLIGENCE, ERRORS AND STRATEGIC CHOICES IN THE REPEATED PRISONERS' DILEMMA

Abstract

A large literature in behavioral economics has emphasized in the last decades the role of individual differences in social preferences (such as trust and altruism) and in influencing behavior in strategic environments. Here we emphasize the role of attention and working memory, and show that social interactions among heterogeneous groups are likely to be mediated by differences in cognitive skills. Our design uses a Repeated Prisoner's Dilemma, and we compare rates of cooperation in groups of subjects grouped according to their IQ, with those in combined groups. While in combined groups we observe higher cooperation rates and profits than in separated groups (with consistent gains among lower IQ subjects and relatively smaller losses for higher IQ subjects), higher IQ subjects become less lenient when they are matched with lower IQ subjects than when they play separately. We argue that this is an instance of a general phenomenon, which we demonstrate in an evolutionary game theory model, where higher IQ among subjects determines -- through better working memory -- a lower frequency of errors in strategy implementation. In our data, we show that players indeed choose less lenient strategies in environments where subjects have higher error rates. The estimations of errors and strategies from the experimental data are consistent with the hypothesis and the predictions of the model.

JEL Classification: N/A

Keywords: Repeated Prisoner's Dilemma, Cooperation, Intelligence, IQ, Strategy, Error in Transition

Eugenio Proto - eugenio.proto@gmail.com
Adam Smith Business School, University of Glasgow and CEPR

Aldo Rustichini - aldo.rustichini@gmail.com
University of Minnesota

Andis Sofianos - andis.sofianos@awi.uni-heidelberg.de
University of Heidelberg

Acknowledgements
The authors thank several colleagues for discussions on this and related research, in particular Pierpaolo Battigalli, Maria Bigoni, Marco Casari, Guillaume Frechette, Drew Fudenberg, David Gill, Marco Lambrecht, Fabio Maccheroni, Salvatore Nunnari, Joerg Oechssler, Emel Ozbay, Erkut Ozbay, Andy Schotter. A discussion of the model with David Levine was very useful. We also thank the conference and seminar participants at the University of Cyprus, University of Zurich, George Mason University, University of Bologna, University of Siena, Bocconi University, Purdue University, University of Maryland and the HeiKaMaX, the 2019 CESifo

Area Conference on Behavioral Economics in Munich. The University of Heidelberg provided funding for this research. AR thanks the National Science Foundation, grant NSF 1728056.

Intelligence, Errors and Strategic Choices in the Repeated Prisoners' Dilemma

Eugenio Proto[*], Aldo Rustichini[†], Andis Sofianos[‡§]

January 21, 2020

Abstract

A large literature in behavioral economics has emphasized in the last decades the role of individual differences in social preferences (such as trust and altruism) and in influencing behavior in strategic environments. Here we emphasize the role of attention and working memory, and show that social interactions among heterogeneous groups are likely to be mediated by differences in cognitive skills. Our design uses a Repeated Prisoner's Dilemma, and we compare rates of cooperation in groups of subjects grouped according to their IQ, with those in combined groups.

While in combined groups we observe higher cooperation rates and profits than in separated groups (with consistent gains among lower IQ subjects and relatively smaller losses for higher IQ subjects), higher IQ subjects become less lenient when they are matched with lower IQ subjects than when they play separately.

We argue that this is an instance of a general phenomenon, which we demonstrate in an evolutionary game theory model, where higher IQ among subjects determines – through better working memory – a lower frequency of errors in strategy implementation. In our data, we show that players indeed choose less lenient strategies in environments where subjects have higher error rates. The estimations of errors and strategies from the experimental data are consistent with the hypothesis and the predictions of the model.

JEL classification: C73, C91, C92, B83
Keywords: Repeated Prisoner's Dilemma, Cooperation,Intelligence, IQ, Strategy, Error in Transition

---

[*] University of Glasgow, 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, Drew Fudenberg, David Gill, Marco Lambrecht, Fabio Maccheroni, Salvatore Nunnari, Joerg Oechssler, Emel Ozbay, Erkut Ozbay, Andy Schotter. A discussion of the model with David Levine was very useful. We also thank the conference and seminar participants at the University of Cyprus, University of Zurich, George Mason University, University of Bologna, University of Siena, Bocconi University, Purdue University, University of Maryland and the HeiKaMaX, the 2019 CESifo Area Conference on Behavioral Economics in Munich. The University of Heidelberg provided funding for this research. AR thanks the National Science Foundation, grant NSF 1728056.

1 Introduction

How do people with different cognitive skills strategically interact in a non-competitive environment? Intelligence is an important characteristic affecting strategic behaviour (e.g. Jones, 2008; Gill and Prowse, 2016; Alaoui and Penta, 2015). Accordingly, in repeated games of cooperation the level of intelligence of players is a crucial factor, Proto et al. (2019) find that when subjects are allocated into two groups on the base of their intelligence, only the higher intelligence groups converge to full cooperation in complex non-zero sum games like the repeated prisoners' dilemma. However, such separation of individuals in distinct classes of intelligence does not occur in everyday life and the question how a group interacts and 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 profitably cooperate more when mixed with the more intelligent. Cooperation rates in the combined treatment increases for the less intelligent players, and slightly decreases for the more intelligent. Specifically, the cooperation rate is substantially higher for lower IQ players (those with IQ in the range 76-106) in the combined treatment than in the split treatment. Instead, the cooperation rate is slightly lower for the higher IQ players (range IQ 102-127), again compared to the split treatment. We identify a critical difference in the frequency of strategies in the combined treatment as compared to the split treatment with higher IQ players. We observe a shift towards 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 increases, and thus the average probability of errors also increases, players 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 are modeled as differences in the working memory. A lower working memory entails a larger probability of error in implementing a strategy. We focus here not on the errors in the choice of action, but on errors in the management of the strategy, i.e. *errors in transition*. We model the strategy as an automaton where 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 and the best response Schlag (1998); Gilboa and Matsui (1991)). The cognitive skills distribution determines the error rates which in turn determine the strategy; high error rates lead subjects to shift towards always defect, the least lenient strategy, which leads to the least efficient outcome.

Furthermore, strictly following the hypothesis of model, we estimate the error rates and the equilibrium strategy and we find a pattern consistent with the predictions of the model, i.e. in environments where subjects commit more errors stricter strategies are observed.

A large literature in behavioral economics has emphasized in the last decades the role of individual differences in attitudes towards others as in theories of social preferences like trust and altruism (facets of Agreeableness of the Big Five) as determining behavior in strategic environments (see e.g. Fehr and Schmidt, 2006, for a survey). Here we explore a completely different mechanism and, to the best of our knowledge, this is the first model in the literature analyzing the relationship between error rates and strategic choices. Error rates are determined by individual cognitive skills, specifically working memory. Bordalo et al. (2017) emphasize the centrality of working memory even in simple economic choices in a recent contribution.

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 choose efficient strategies leading to cooperation under repeated interactions. Fudenberg et al. (2012) analyse the effect of exogenously induced 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. A key difference with respect to our setting is that the subjects know that their action can be implemented with an exogenously induced error.

There is evidence of teaching in social experimental environments. Subjects actively teach other subjects how to play efficiently in the laboratory (e.g. Camerer et al., 2002; Hyndman et al., 2012; Cason et al., 2013). In our setting there is no active teaching, individuals update their beliefs by observing partners' choices which affect their behaviour in the next interaction. Some recent findings indicate that lower intelligence is associated with more reliance on social learning through imitation (e.g. Vostroknutov et al., 2018). In this literature individuals are generally aware of their difference, while in our design they are not directly observable.

The paper is organized as follows: In section 2 we present the experimental design. In section 3, we show the experimental evidence. In section 4 we describe the main results of a model of evolutionary game theory that we analyze in detail in section 5. In section 6, we estimate the main parameters and variables of the model using our experimental data where we find support for its main predictions. In section 7, we discuss some of the assumptions made in the model and we conclude in section 8. Technical analysis of the model and details about its estimation, robustness checks, further details of the experimental design and descriptive statistics are in the appendix.

2 Experiment

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 randomly determined 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.[1]

[1] 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 design

Day One

On the first day of the experiment, participants are asked to complete a Raven Advanced Progressive Matrices (APM) test consisting of a of sequence of 36 questions. They have a maximum of 30 minutes for the entire test. For each item a 3×3 matrix of images is displayed on the subjects' screen; the image in the bottom right corner is missing. The subjects are then asked to complete the pattern choosing one out of 8 possible choices presented on the screen. Before initiating the test, the subjects are shown an example of a matrix with the correct answer provided below for 30 seconds. The 36 questions are presented in order of progressive difficulty as they are sequenced in Set II of the APM. Participants are allowed to switch back and forth through the 36 questions during the 30 minutes, and are allowed to 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 Gill and Prowse (2016) and Proto et al. (2019) and has been found to be relevant in determining behaviour in cooperative or coordinating games.

In psychometric, and also experimental economics research, subjects are usually not rewarded for completing IQ tests such as the Raven. However, it has 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 mention that this is a test of intelligence or cognitive abilities.

Following the administration of the Raven test, participants complete an incentivised Holt-Laury task (Holt and Laury, 2002) to measure risk attitudes. Finally, participants are 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 is offered for this section of the session and the subjects are informed about this. We use the Big Five Inventory (BFI); the inventory is based on 44 questions with answers coded on a Likert scale. The version we use 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.[2]

Day Two

On the second day, the participants are asked to come back to the lab and are allocated to two separate experimental sessions. The basis of allocation depends on the treatment. In the IQ-split

[2]This is available online at `https://drive.google.com/open?id=13wL3CwP1nqZ3b84om81OfzAFyJOz-1Py`

treatment, participants are invited back according to their Raven scores: subjects with a score higher than the median are 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.[3,4] In the combined treatment, we make sure to create groups of similar Raven scores across sessions. To allocate participants to second day sessions, we rank them by their Raven scores and split by median. Instead of having high- and low-IQ groups though, we alternate in allocating participants in one session or the other.[5]

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

We induce 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 is $\delta = 0.75$. We use 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 is either 45 minutes or until the 151st round is played depending on which comes first.

The game parameters 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.[6]

The matching of partners is done within each session under an anonymous and random rematching protocol. Partipicants play as partners for as long as the random continuation rule determines that the particular partnership is to continue. Once each match is terminated, the subjects are again randomly and anonymously matched and start playing the game again according to the respective continuation probability. Each decision round for the game is complete when every participant has made their decision. After all participants make their decisions, a screen appears that reminds them of their own decision, indicates their partner's decision while also indicating the units they earned for that particular round. The group size of different sessions varies depending on the numbers recruited in each week.[7] The participants are paid the full sum of points they earn through all rounds of the game. Payoffs reported in table 1 are in terms of experimental units; each experimental unit corresponds to 0.003 Euros.

Upon completing the PD game, the participants are asked to respond to a short questionnaire

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

[4] 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-IQ session.

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

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

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

about any knowledge they have 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 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 online.[8]

3 Experimental Evidence

3.1 Cooperation rates and 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 their choice of first move across all treatments. Subjects in the high-IQ sessions converge faster to almost total cooperation rates, 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). This result replicates the findings in Proto et al. (2019) by using a different subject pool in a different country. Subjects in the combined session have higher cooperation rates than in the low-IQ sessions and lower than the ones in the high-IQ sessions (right panel).[9] Table 2 shows that in the high-IQ sessions subjects earn about 2.5 units and cooperate about 10% more than in the combined sessions in the first 20 supergames, while in low-IQ sessions they cooperate about 22% and earn about 5.6 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 the less intelligent learn to play as efficiently as the more intelligent in the second part of the session in the combined treatment. Meanwhile in the low-IQ sessions the difference in both cooperation and payoffs remains constant. In accordance to the findings in Proto et al. (2019), 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 difference in coopera-

[8]See note 2.

[9]In figure A.2 of the appendix we present the cooperation rates by session.

tion between individuals of different cognitive skills is only due to some learning effect during the session.[10]

Figure 2 shows that the average payoff per interaction is consistently higher in the combined sessions than in the high-IQ and low-IQ split sessions, indicating that in the combined treatment subjects play on average more efficiently than in the split treatments.[11]

Finally, we can summarize the findings in this section by explicitly stating that *cooperation rates and aggregate payoffs are higher when high and low IQ subjects are combined together than when they play separately*

3.2 Learning

Table 4 shows that – in the first 20 supergames– subjects in the high-IQ sessions increasingly open with cooperation earlier than in the combined treatment, while in the low-IQ session the cooperation increase is later 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 the less intelligent learn to play more cooperative strategies when mixed with the more intelligent faster than when they play together.

Why does this earlier increase in first periods cooperation in the session with a larger number of high IQ subjects occur? The increase in cooperation seems, at least in part, to be driven by an increase of subjects' beliefs about the way the other subjects open with. In columns 2 and 4 of table 4 we can observe that the coefficients of the partners' past cooperation in the 1st periods are positive and significant. In other words, subjects whose partners opened with cooperation in the past supergames are more likely to open with cooperation in the present.

3.3 Choice of strategies in the different environments

There is widespread evidence that subjects overwhelmingly play *memory one* strategies in the repeated prisoners' dilemma game (e.g. Dal Bó and Fréchette, 2018; Proto et al., 2019). Accordingly, the choices in every round of a supergame are determined by the past outcome (hence by both own and partner's choice in the previous period of the supergame). Let then $ch_{i,t}$ represent the subjects' choice (1 for *cooperate* and 0 for *defect*), $Partn.Ch_{i,t}$ represent partner's choice and $p_{i,t}$ represent the probability of $ch_{i,t} = 1$ (conditioned on the set of independent variables), we then have the following model:

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

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 are the time-invariant

[10]Interestingly, risk aversion is the only significant determinant of cooperation at the beginning of each session.

[11]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.

individual fixed-effects (taking into account time-invariant characteristics of both individuals and sessions); finally $\epsilon_{i,t}$ is the error term. We estimate model 1 separately for the first 20 and the second 20 supergames by using a logit estimator.

In table 5, we present the estimates of model 1 for the 1st and the 2nd block of supergames. Results are reported in odds ratios and taking the outcome $(C,C)_{t-1}$ as baseline. In panel A, we note that the odds of cooperating at time t by high-IQ are higher when they play among themselves than when they play in the combined treatment conditional on deviation from mutual cooperation (i.e. after $(D,D)_{t-1}$ $(D,C)_{t-1}$ $(C,D)_{t-1}$). This difference is, if anything, even larger in the second part of the session as seen in panel B of table 5.[12] This suggests that high-IQ revert to cooperation less frequently, hence that they adopt less lenient strategies when playing in the combined sessions than in the split sessions. In table 6 we report a direct test on whether high-IQ are more forgiving when they interact amongst each other than when they are in the combined treatment. We note that the high IQ are significantly less likely to cooperate whenever the other subject unilaterally defects. The low-IQ do not seem to play significantly differently whether they play with with other low-IQ or in the combined treatment after an unilateral defection, but the sign of the coefficient is positive. Hence we can summarise this section by saying that *high-IQ are significantly less likely to cooperate after a unilateral deviation of the partner when they interact in the combined treatment than when are in the split treatment.*

This preliminary evidence shows that subjects in general learn to cooperate and choose their strategies on the basis of their cognitive skills, but also on the distribution of the cognitive skills within the group. The high-IQ seem to be more forgiving when they play with other high-IQ players than when they play in the combined treatment. Therefore, a complex mechanism is in place and the following model allows to analyze this in detail.

4 Errors and Strategy Evolution: Summary

We model differences among players or among groups as differences in working memory. A lower working memory entails a larger probability of error in implementing a strategy. We focus here not on the errors in choice of action, but on errors in the management of the strategy, and define this as *error in transition*, (for more details see section 5). 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 will result in 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 assume that subjects only play Always Defect, Tit For Tat and Grim Trigger strategies, following Dal Bó and Fréchette (2019), who show that subjects mostly utilise these strategies. To lighten notation, in the current and next section we denote these strategies as $\{A, G, T\}$, where A

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

stands for *Always Defect* strategy, G for *Grim Trigger* and T for *Tit-for-Tat*.[13]

The main results of the analysis (developed in detail in section 5) are the following:

First, the limit behavior of the fraction of strategies is determinate (that is, the steady states are locally unique), if, and only if, there are errors. When players have perfect working memory, and there are no errors, neither of action or of transition, equilibria of the strategy choice game, and steady states of the learning process and the evolutionary model are not locally unique (with the exception of the (A, A) equilibrium, which is locally unique for all values of the parameters). Instead, when errors occur, even of arbitrarily small size, 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.

Second, 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 a function of the error rate. The size of these basins changes with the size of the errors in a natural way. As ϵ becomes larger (that is, as more error prone the group of players is), the basin of attraction of 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.

Third, which strategies survive in the long run depends on δ; in all cases, for low rates of errors all strategies may survive depending on the initial condition. For high rates of error only defection survives. For intermediate rates, the two surviving strategies are defect and grim trigger for low δ, and defection and Tit-for-Tat for high δ (the details are in Conclusion 5.13 of section 5). As δ tends to 1, keeping error rates fixed, 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.

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. In section (G) 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, specifically the best response dynamics.

4.1 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 3. This figure reports the phase portrait of the vector field for the Best Response dynamics, at different error rates, ranging from 0 (no error, top left panel) to 0.25 (bottom right panel). The top right panel shows the case with a very small

[13] Dal Bó and Fréchette (2011) and Proto et al. (2019) find evidence that these are found to be between 66% and 90% in likelihood of being played by their participants.

error rate (0.0001). The two top panels of the figure illustrate the sudden change of the dynamic as soon as the error becomes positive but very small. Payoffs in the stage game are set as in our experimental design in table 1. The discount factor (or equivalently, the continuation probability) is the same as in our experimental design, namely 0.75.

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.[14] 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 two 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.9, reporting the same results for the Replicator Dynamic (or Proportional Imitation).

Several points that appear clearly in Figure 3 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 are considering in this example. 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 most 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.*

4.2 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 that the A strategy is for every error probability the unique best response to itself (see Lemma 5.5). 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

[14] 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 (3), the distinction between differential inclusion and differential equation is irrelevant.

payoffs (which is satisfied by the payoffs in the experimental design we adopt) the (A, A) is the unique Nash (see Lemma 5.6). These two facts together with the continuity of the value function in ϵ (see Lemma 5.3) 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 rate is made possible, the value for (say) player 1 of using G with G (call it $V(GG)$) compared to the similarly defined values $V(TG)$, $V(GT)$, $V(TT)$ 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 no meaning for basin of attraction of G and T. 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 section 5). 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 the 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.

4.3 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 large δ, close to 1. Consider players starting from mutual cooperation. A small transition error can lead one to defect. 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 T increases fitness, and its basin of attraction is larger.

The increase in the error rate (modelled 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). For a graphical illustration of the role of the discount factor, compare Figure 3 (where $\delta = 0.75$) with Figure A.8 (where $\delta = 0.6$); in both cases payoffs are the same as in our experimental design.

4.4 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 two different models to argue this point: 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). We describe both models in section 5 below. We will then describe the best response dynamics in detail and simulate the resulting evolution of cooperation among individuals belonging to different intelligence groups and treatments following Dal Bó and Fréchette (2011) in the section G of the appendix. The simulation approximates quite well the dynamics of cooperation observed in our data (figure A.12 in the appendix), the estimation also emphasises some interesting differences in the beliefs updating according to IQ and the treatment a subject is within.

5 Errors and Strategy Evolution: Technical Details

In this section we present a more detailed exposition of the model of strategy evolution when players can make errors. Readers who are satisfied with the summary presented in section 4 can omit reading this section with no substantial loss.

Subjects can make two types of errors when implementing their strategies. Firstly, errors 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 do in the future. We represent this here (through modelling 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.

We focus in this section on the simple case of only transition errors. In section F of the appendix we consider the general case in which both errors in action choice and in transition are possible. The reduction to transition error considerably simplifies the exposition, and the results in section F show that this case contains all the essential features. Our data analysis testing the model estimates the complete model with both types of errors. The method of the estimation is presented in section H of the appendix. Finally in section G of the appendix we present a model of population learning which allows us to establish a link with social learning models as the one used in Dal Bó and Fréchette (2011).

In the study of the model we will investigate the effect of error rates on the frequency of strategies. It is useful to have in mind the order of magnitude of error rates that is actually observed in our data. A summary estimate of the frequency of error rates is provided in section 6.1. In figures A.4 and A.5 we report the average error rates in transition and action. The dispersion of transition error rates across individuals is reported in figures A.6 (for the early supergames) and A.7. The range of error rates, in the experimental groups we consider, is for almost all participants within the range of 0 to 0.3. The mean is between 0.2 and 0.01, but figure 3 shows that in earlier supergames the mean can be much higher, around 0.15 to 0.25. Overall, it seems reasonable to focus our attention conservatively (to consider all possible relevant cases) to the range 0 to 0.25.

Setup

The stage game is a Repeated Prisoner's Dilemma game with payoff:

	C^2	D^2
C^1	c, c	s, t
D^1	t, s	d, d

where

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

These conditions are satisfied by the payoffs in our experimental design, with a strict inequality for the last. We focus on the set of the most commonly used repeated game strategies. We consider in the following an 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. 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. 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. Finally, P defines the transition probability where

$$P(\cdot; x, (a^1, a^2)) \in \Delta(X) \tag{3}$$

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

In the analysis that follows, when we introduce errors in transition, we will consider the automata G_C and G_D as 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.

5.1 Equilibrium and Evolutionary Dynamics with No Errors

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

$$M \equiv \{A, G, T\} \tag{4}$$

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

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

We denote by μ_R the strategies in the reduced game.

By assumption (2), $d > (1-\delta)s+\delta d$. 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

$$\delta^* \equiv \frac{t-c}{t-d}. \tag{5}$$

We consider in the following only the case in which

$$\delta > \delta^* \tag{6}$$

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

$$\mu_R(A,\delta)^* \equiv \mu^* = \frac{c-(1-\delta)t-\delta d}{c-(1-\delta)t-(1-\delta)s-\delta d+(1-\delta)d} \tag{7}$$

Proposition 5.1. *When (6) 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. 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

$$\forall m \in M, \frac{d\mu(t,m)}{dt} = \mu(t,m) \left(U(m,\mu) - U(\mu,\mu) \right) \tag{8}$$

The best response correspondence is defined as taking values in mixed strategies:

$$BR(\mu) \equiv \{ \tau \in \Delta(M) : \forall m \in M, U(\tau, \mu) \geq U(m, \mu) \} \tag{9}$$

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

$$\forall m \in M : \frac{d\mu(t,m)}{dt} \in BR(\mu(t, \cdot))(m) - \mu(t,m). \tag{10}$$

To clarify the notation, we note that $\mu(t, \cdot)$ is a probability measure, BR is the best response correspondence applied to μ, and $BR(\mu(t, \cdot))(m)$ is the probability assigned by the best response to m; so $BR(\mu(t, \cdot))(m)$ is a set. We indicate mixed strategies or frequencies in the following as a vector $(\mu(A), \mu(G), \mu(T))$.

The justification of (8) is standard (Sandholm (2007)). A justification of the equation (10) within a population model is provided in appendix G.1.1. In this version of the model, players in a large population know exactly, and at every point in time, the frequency of strategies in the population, but, by assumption, only a fraction have the opportunity to revise their current strategy. When they do, they adopt the best response to the current distribution. The best response dynamics is one step closer to a model of rational learning, so it is a useful intermediate step to build our intuition regarding it. It differs from the rational learning model because it requires that all players have the same belief and the belief is correct. In addition, as we are going to see (proposition 5.2), the qualitative behavior of the strategy frequency is close to that predicted by the proportional imitation model; so the study of the best response dynamics shows the robustness of the results to changes in the detail of the dynamic model.

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 a small change in the initial conditions alters the long run behavior. More precisely:

Proposition 5.2. *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*

$$\{ (\mu^*, \mu^* p, \mu^*(1-p)) : p \in [0, 1] \} \tag{11}$$

and the interval

$$\{(0, p, (1-p)) : p \in [0,1]\}; \tag{12}$$

2. *the path with initial condition* $\mu(0, A) > \mu^*$ *converge to the steady state* $(1, 0, 0)$*; the paths with initial conditions where* $\mu(0, A) < \mu^*$ *converge to a steady state in the set (12) above;*

3. *The steady states in (11) are all unstable.*

In the proportional imitation dynamic, 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*

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

In conclusion, evolutionary dynamics 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 will show in the analysis below that the frequency is precisely determined in the long run when an error of arbitrarily small size is possible. Figure (5) below illustrates the dynamic behavior of the proportional imitation model when no errors occur. The dynamic behavior with Best Response dynamics is very similar (see figure 3), but in such degenerate cases (in which the payoffs for the strategies G and T are the same) there is a large multiplicity of solutions of the differential inclusion, and the dynamic is very sensitive to the tie-breaking rule.

The dynamics are 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.

5.2 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, (e.g. they 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 this parameter, which describes an individual characteristic: the higher the intelligence, the lower the ϵ.

To compute the payoff from the choice of the strategy when transition errors 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:

$$S \equiv \{A, G_c, G_d, T_c, T_d\} \tag{13}$$

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 $s \in S$, 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:

$$\Omega \equiv S \times S \tag{14}$$

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

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).[15]

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

$$\forall \omega = (s^1, s^2), \omega' = (r^1, r^2), \text{ if } a(\omega) \equiv (a^1(s^1), a^2(s^2)),$$

$$Q(\omega'; \omega) \equiv P(r^1; s^1, a(\omega)) P(r^2; s^2, a(\omega)) \tag{15}$$

We then define the error transition as the stochastic matrix $E : S \to \Delta(S)$ sending each automaton of a type to the same type automaton, but choosing the state with $1/2$ probability.[16] We denote

[15]More precisely, we can write this transition for player 1's automaton:

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

[16]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. $\forall a : E(A; A, a) = 1.$

by $q^i(\omega)$ 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:

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

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

$$
u(\omega) = u(s^1, s^2) = R(a^1(s^1), a^2(s^2)) \tag{16}
$$

5.3 Value Function

Players choose an element in the set $\{A, G, T\}$, 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 the value function $V : \Omega \to \mathbb{R}$.

Lemma 5.3. *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:

$$
V = (1-\delta)u + \delta Q_\epsilon V \tag{17}
$$

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

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

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.

It is clear that some 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

$$\mathcal{P}(\Omega) \equiv \{\Omega_{AG}, \Omega_{AA}, \Omega_{AT}, \Omega_{GA}, \Omega_{GG}, \Omega_{GT}, \Omega_{TA}, \Omega_{TG}, \Omega_{TT}\}. \tag{18}$$

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 equation (17) with (u, Q_ϵ) replaced by $(u_i, Q_{\epsilon,i})$; these equations can be solved and analyzed independently.

Lemma 5.4. *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. $V(GA) = V(TA) = s(1 - \delta(1 - \epsilon)) + d\delta(1 - \epsilon)$

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 (5.5) 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 5.5. *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

	G_cA	G_dA
G_cA	ϵ	$1-\epsilon$
G_dA	ϵ	$1-\epsilon$

Using equation (17), we can solve for $V(G_cA)$ and find:

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

Therefore for all $\epsilon > 0$,

$$V(G_cA) < d = V(AA). \tag{19}$$

Since the transition matrix restricted to Ω_{AT} is the same as the one we reported for Ω_{AG}, we conclude:

$$V(T_cA) < V(AA). \tag{20}$$

□

5.3.1 Uniform Errors

Using invariant sets 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.[17]

Lemma 5.6. *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

$$m_s \equiv (1-\delta)s + \delta\frac{s+d}{2}, m_t \equiv (1-\delta)t + \delta\frac{t+d}{2}, \tag{21}$$

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

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

Proof. For any element $i \in \mathcal{P}(\Omega)$ (this set is defined in (18)), we have that the transition restricted to the point in i is:

$$Q_{(1/2),i} = \frac{1}{\#i}U_i \tag{22}$$

where U_i is a square matrix of 1's of dimension $\#i$ (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:

$$V_i = (1-\delta)u_i + \delta M_i V_i \tag{23}$$

Equation (23) implies

$$M_i V_i = M_i u_i \tag{24}$$

Now (23) and (24) give the formula for the value in terms of the payoffs:

$$V_i = (1-\delta)u_i + \delta M_i u_i \tag{25}$$

[17] We consider here the case in which no transition error is possible when no transition has occurred yet, that is at the first round of the game. The case in which an automaton is chosen and an error in the choice of the initial state is made (which may be considered as a transition error) the conclusion of the lemma (5.6) is even easier to reach.

The rest follows from simple algebra.

For the last statement, from our assumption (2) on the stage game payoffs we conclude that $m_t > m_c$. We already know from lemma (5.5) that $d > m_s$ holds for any δ. Thus A is a dominant action. □

5.3.2 Small Errors

We know that for "large" errors ($\epsilon = 1/2$) 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:

$$\bar{\epsilon} \equiv \sup\{\epsilon : V(GG) > \max\{V(TG)), V(AG)\} \tag{26}$$

$$\& V(TT) > \max\{V(AT)), V(GT)\}\}$$

Lemma 5.7. *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 (2) for which $\bar{\epsilon}$ is finite.*

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

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

where the strict inequality follows from our assumption (6). Thus we conclude that $\bar{\epsilon} > 0$ from lemma (5.3). The second claim also follows from lemma (5.3). From last claim follows from lemma (5.6), which gives as simple sufficient condition for $\bar{\epsilon}$ to be finite. □

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

Lemma 5.8. *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 side* $\mu(G) = 0$ *(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:*
$$\mu(G)^* = \frac{V_\epsilon(TT) - V_\epsilon(GT)}{V_\epsilon(TT) - V_\epsilon(GT) + V_\epsilon(GG) - V_\epsilon(TG)} \tag{27}$$
where V_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

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

As $\epsilon \to 0$ both numerator and denominator in the right hand side of (28) tend to zero. Now the conclusion follows from l'Hôpital rule. □

With small errors, 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.

5.4 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:
$$A \succ G \succ T \tag{29}$$

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 5 and 3. 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 $s_{ij}(\epsilon, \delta, u)$.

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

$$\mu(A) = \frac{G(B)}{G(A) + G(B)}. \tag{30}$$

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 (SBA) of A is:

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

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.

5.5 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 $\epsilon = 1/2$:

Proposition 5.9. *Under proportional imitation,*

$$s_{AG}(0, \delta, u) = s_{AT}(0, \delta, u) = \frac{(1-\delta)d - (1-\delta)s}{c - (1-\delta)t - (1-\delta)s - \delta d + (1-\delta)d} \tag{32}$$

$$s_{AG}(1/2, \delta, u) = s_{AT}(1/2, \delta, u) = s_{GT}(1/2, \delta, u) = 1 \tag{33}$$

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.

5.6 Payoffs in the Experiment

We now analyze the basin of attraction for the specific numerical values of the payoffs used in the experiment. In section D of the Appendix we report numerical values of the payoff matrix with errors and figures portraying the basin of attraction. To see how the size of the basin of attraction changes when the stage game payoffs are those we used in the experiment ($c = 48$, $s = 12$, $t = 50$, $d = 25$), we can use the lemmas (5.3) and (5.4) to compute the value function and analyze equilibria and dynamic behavior.

We refer to figure 4, which describes the change for different values of ϵ (moving along the x-axis in each panel) and δ.

The first conclusion is:

Conclusion 5.10. *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 5.11. *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 5.12. *The lenient strategy T can survive only with low errors or with high probability of continuation.*

Figures (6) illustrates the dynamics of the proportional imitation for values above and below the threshold, and payoffs 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 5.13. *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; in the intermediate region the attractors are A and G.*

With small errors, 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.

Also with higher errors 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.

5.7 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 $\{A, G, T\}$, 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 $\epsilon = 0.35$ ($\delta = 0.9$) (values are reported in table A.16) 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.16, the pure strategy G is a steady state of the PID, but is not a Nash equilibrium of the complete game, because $V(AG) > V(GG)$.

Proposition 5.14. *With best response dynamics,*

1. *for $\epsilon < \bar{\epsilon}$, 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 (5.10), (5.11), (5.12) and (5.13) hold in the case of BRD as well; in particular, for $\epsilon > \bar{\epsilon}$ 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 similar. 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 3 that we have already discussed in our introductory discussion in section 4 and we refer the reader to that section.

6 Errors and strategic choices in our data

In the two sections above we saw that, as error rates increase, subjects play more often stricter strategies. If we assume that error probability (in particular the probability of transition error associated with working memory) is negatively correlated with cognitive abilities, we should observe stricter strategies among the subjects in the low-IQ split treatment if compared to the combined treatment, and in the combined treatment when compared to the high-IQ split treatment. In this section we show that our experimental data fit the assumption and the implied predictions of the model.

In the section H of the appendix we describe in detail the algorithm we use to identify errors and the strategy played (from the three strategies we restrict in the model). Table A.20 in the appendix shows that the error in transition individual rate is significantly higher in the low-IQ group in the split treatment, and in the combined treatment, when we compare them to the high-IQ group in the split treatment. The error in transition rate is also negatively correlated with individual IQs. The correlation between errors in action and IQ seem less clear, as we would expect according to the hypothesis that errors in transition are working memory errors.

6.1 Error Rates and Strategy Frequency

In table 7 we present a summary of the results of these estimations: the frequency of the strategies played in the different treatments and within groups in the combined treatment, the error in transition and error in action rates. Error frequencies – estimating the error rates ϵ and η in the model – are higher, the lower is the proportion of high IQ subjects in the treatment. As predicted in our model the strategies become stricter in environments with higher error rates. The frequency of A is higher in the low-IQ group, split treatment, than in the combined and it is the lowest in the high-IQ split treatment. The same is true for the ratios A/T and A/G.

The ratio G/T does not seem to follow the same pattern of strictly increasing with the errors: in fact it changes very little across the three treatments. Note that for the range of errors we are considering (i.e. less or equal to 0.01) the proportion of the basins of attraction between Grim Trigger and Tit for Tat do not change much in the model. This can be already observed from the two top panels of figure 3 comparing the size of the two basins of attractions of the two strategies. From figure 4 – where we can directly observe the ratios of the size the basins of attraction of the strict strategies– it is clear that the ratio of the basin of G over the basin of T is non monotonic when errors are small.

Figures A.4 and A.5 in the appendix report the average error rates in transition and action. Information on the dispersion of error rates across individuals is provided in figures A.6 (for the early supergames) and A.7. The range of the error rates in the four groups for the most part is within the range 0 to 0.30.

7 Learning Error Avoidance

The model described in the two previous sections assumes that the probability of both errors, ϵ and η, are fixed during each entire session. In our data, the error rates decrease as the subjects gain experience (as suggested by the regression reported in table A.20 in the appendix). In table 8 we analyze the first and the last 10 supergames of each treatment separately. Both ϵ and η decrease in the last part of all treatments compared to the first part. We can observe between periods a similar pattern to the one we observe across treatments: the frequency of A and the ratios A/T and A/G decrease with the errors. Similarly as before, the ratio G/T does not follow this pattern, but –as we already argued– this can be explained by figure 4, where we can observe that the ratio of the basins of G and T is not monotonic especially within the range of errors we are considering (i.e. between 0.007 and 0.005). We discuss below (in section 7.1) the implications of the fact that errors are declining. Finally, note that we observe the same pattern in terms of strategies by using the estimation method of Dal Bó and Fréchette (2011), presented in table A.21 of the appendix.

The estimation of errors and strategies in the section above shows that our data follows the predictions of the model, but also reveal more complexity in the evolutions and the structure of the errors. We will now argue that this complexity does not change the qualitative prediction of the model.

In table 8 we see that error rates towards the end of the session are lower than in the beginning, for each group; so participants during the session learn to play the game. In figure A.3 in the appendix we can observe this pattern in more detail. Both transition and action error rates decline for everyone. They all converge to nearly zero. In the low-IQ split treatment the convergence is slower than in the combined treatment which in turn it is slower than in the high-IQ split treatment, this suggests a process of learning by doing where all subjects adapt to their respective environment. The errors are relatively low at the beginning of the low IQ split treatment, this is probably due to the fact that in the first 10 supergames the share of subjects choosing the always defect strategy is quite high – as we can observe in table 8– and the errors in transition are by definition not possible while playing the always defect strategy, while committing error in action for this strategy is in reality probably less likely.

7.1 Evolution with Learning Error Avoidance

In our model, for simplicity, we assume that the probability of errors, η and ϵ, are fixed. Considering their decline as exogenous processes of learning by doing will not qualitatively change the model's conclusions. Therefore, to get a complete picture of the evolution of strategy frequencies when we apply the analysis to our experimental data, one has to adjust the conclusions to the more general case in which the error rates change over time. The analysis in sections 4 and 5 provides a general guidance.

Consider the entire range of error rates that we observe in our data. The average change is in the range of 0.03 to 0 (see figures A.4 and A.5). Consider first the region in the simplex obtained as intersection of the basins of attraction for a strategy, say always defect, A, over the error rates in that range. Do then the same for the two other strategies. When we take the collection of these subsets of the simplex, we obtain all together a (likely strict) subset of the simplex. On this subset we can make precise predictions: when the initial condition is in one of these sub-regions (say the one for A), then the time path of frequencies will converge to A even if the error rates change. When an initial condition is not in any of the sub-regions obtained in this way, then the analysis depends on the two speeds of adjustment, one of the strategy frequency and the other of the error rates. In summary, the long run behavior is ultimately determined by the limit values of the error rates, which in our data are very low. However, starting with a higher error rate can make the set of initial conditions on the fraction of strategies, that eventually converge to A, larger, because of the initial movement in the direction of A.

Figure 7 reports the evolution of frequency of strategies over time in the different treatments: high-IQ split, low-IQ split and combined. The frequency of A declines in all treatments. The total frequency of grim trigger, G, and tit for tat, T, converges to almost 100 per cent in the high IQ-split treatment, and in the combined treatment. The fraction of A in the low IQ-split treatment seems to stabilize to around a level of 0.20 in the long run. However in the low-IQ group, A is on a higher level (around 0.40) after 5 supergames. As we argued in the preceding paragraph, the initial and transitory levels of error frequency can affect the long-run trajectory. In fact, an initial high level of error can bring the fraction of strategies in the early part of the session in a region close to high values of A, thus in the basin of attraction of A. The ensuing decline of error rates may not be able to revert this initial shift in the direction of the defect strategy

8 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 occurs 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, a natural question arises: what are the outcomes of strategic interactions among heterogeneous individuals? We have presented three main results.

The first is that cooperation rates in heterogeneous groups are closer to the higher cooperation rates that occur within segregated groups of high intelligence, although the more intelligent make a small loss. The entire aggregated surplus is higher when heterogeneous groups play 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 combined groups is due to the influence of the choices of higher intelligence players, who are more consistent in playing optimal strategies, hence commit less errors.

The third and final result concerns the observed shift, in groups of lower intelligence and so with higher error rates, towards the direction of harsher strategies. We argue that this is an instance of a general phenomenon which can be explained by models of social learning. To analyze the relation between error rates and distribution of strategies we proposed an evolutionary game theory model where a population of players who play a sequence of repeated games and choose at the beginning of every supergame a repeated game strategy, in the set of always defect (A), grim trigger (G) and tit-for-tat (T). Populations may differ in the error rate of the implementation of a strategy, either error in the choice of action or in the transition among states of the corresponding automata. Studying the long run distribution of strategies and size of their basins of attraction, we show that players choose more lenient strategies in environments where subjects commit few mistakes (the basin of attraction includes in the limit of small errors only T), and instead shift to stricter strategies when the error rates increase, until only the equilibrium where the entire population plays A survives. We show that the same results, and a similar intuition, operate in different evolutionary models and dynamics, for example replicator or best response dynamics. The analysis of the data, testing the model, supports the main assumption (that error rates are negatively correlated with intelligence) and the main predictions on how the distribution of strategies depends on the error rate.

These results provide a useful guidance for policies generating social interactions among groups heterogeneous in income or education level such as the one fostered by Moving to Opportunity (MTO) policies. In particular, we show that social interactions are likely to be mediated by differences in cognitive skills. Hence, our design can be considered a controlled MTO, where we can compare rates of cooperation and differences in strategies in groups segregated according to cognitive skills with those in mixed groups.

An important extension of these results should examine whether and how these results extend (in experiments and in theory) to general games, and not just in the Repeated Prisoner's Dilemma (RPD), and for more comprehensive sets of repeated game strategies. In the RPD, the association between smaller error rates and more lenient strategies follows from the way in which the value function, which associates payoffs to strategies, changes with the error rate. For example, T and G give the same payoff (and so are equivalent) in an evolutionary model with no errors. For a similar reason these two strategies are equivalent in a rational learning model with no active experimentation. This conclusion changes dramatically even for very small errors, because the loss of payoff for G is higher than the loss for T, hence the result that the basin of attraction of T is larger. With stage games in the repeated game different from the Prisoner's dilemma the value function associating payoffs to errors is easy to determine. But how this affects the relative fitness properties, and thus the predicted frequency of different strategies in evolutionary models, is an open problem. The same is true for models of rational learning with experimentation and errors. This seems to be an important question if we want to understand how game theoretic predictions extend to a world where players make mistakes, perhaps with a frequency associated with the level of their cognitive skills.

Another important progress might be provided by a direct test of the evolutionary or rational model. This could be accomplished with an experimental design providing appropriate measurement and control of the belief process. Such a study would require belief elicitation from participants along the session, either by direct observation (for example using eye-tracking measurements) or by surveys. Also, initial conditions on beliefs could be manipulated, to insure some control over the initial condition on beliefs. This could be accomplished for instance by providing subjects at the begging of the session with some information on the behavior of participants in a previous sessions, and checking with belief elicitation whether and how much this manipulation has been effective. This is topic for current and future research.

9 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 low IQ split sessions are the the black and grey lines respectively in the left panel. Combined treatment sessions correspond to the middle grey lines in the right panel.

.

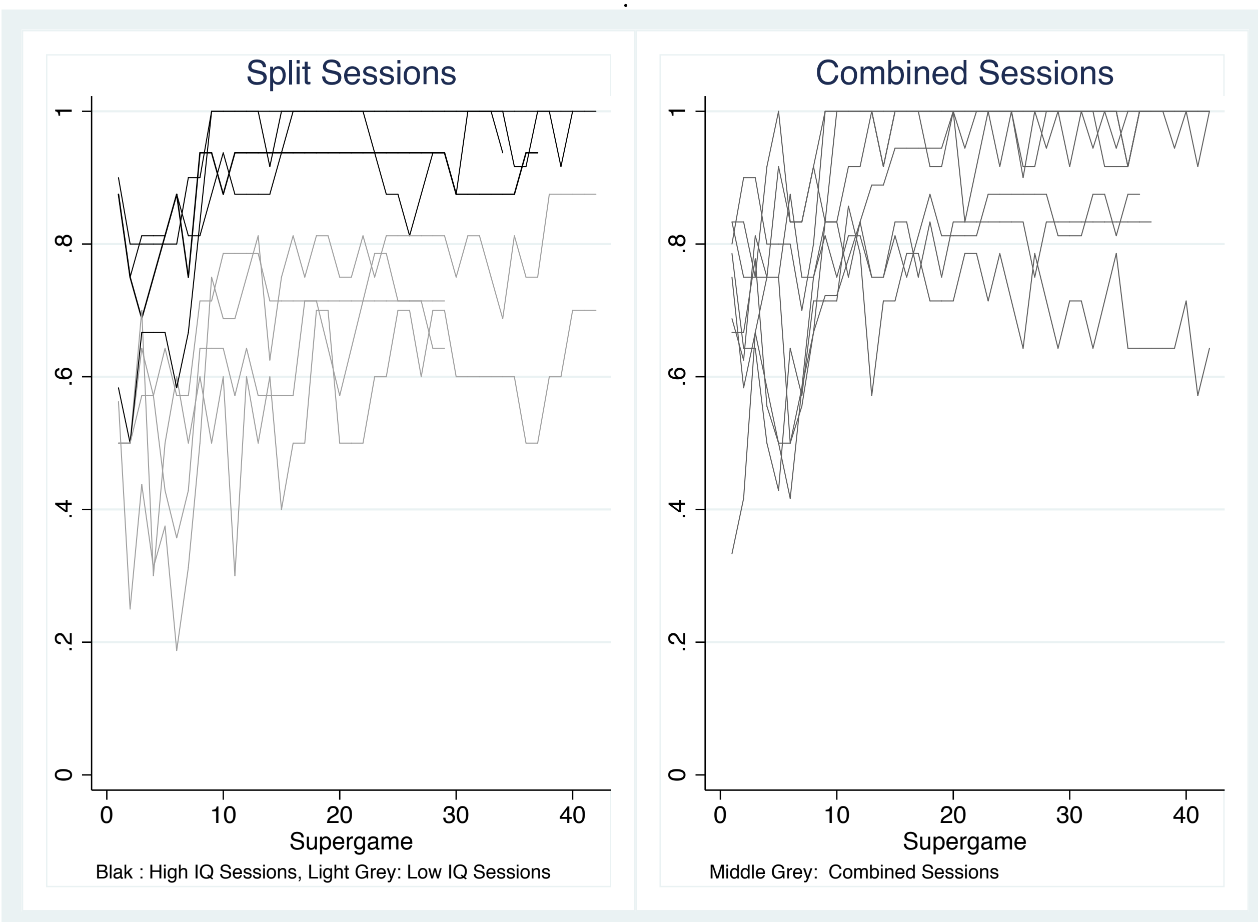
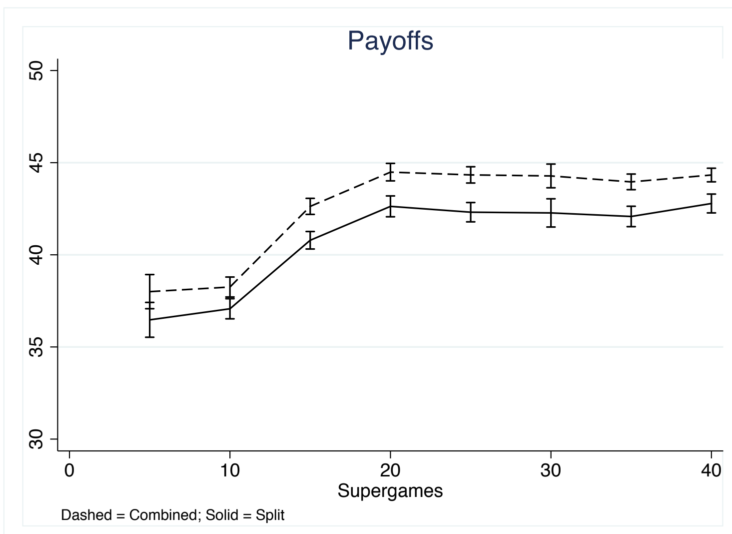


Figure 2: **Average payoffs per interaction in the Split and Combined sessions.** The average is computed over observations in successive blocks of five supergames, of all Split and Combined sessions, aggregated separately. Bands represent 95% confidence intervals.



.

Figure 3: **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 0 to 0.25. Payoff and discount factor ($\delta = 0.75$) are as in our experimental design.

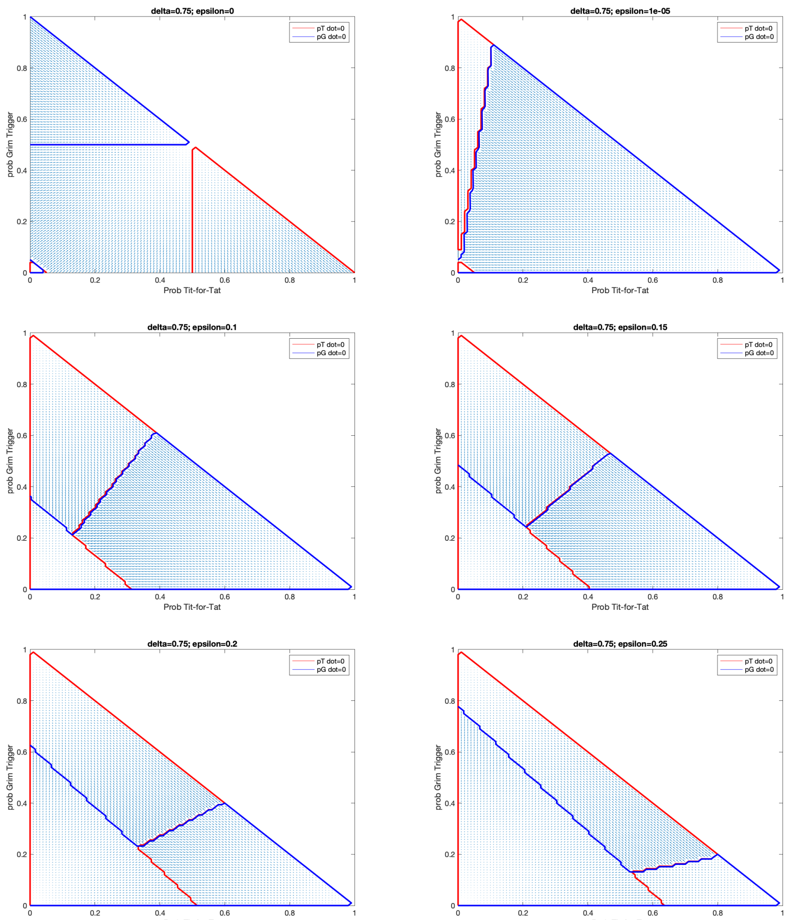


Figure 4: **Size of basin of attraction of strict strategies.** Top to bottom panel: values of δ equal to 0.75, 0.82, 0.9 respectively.

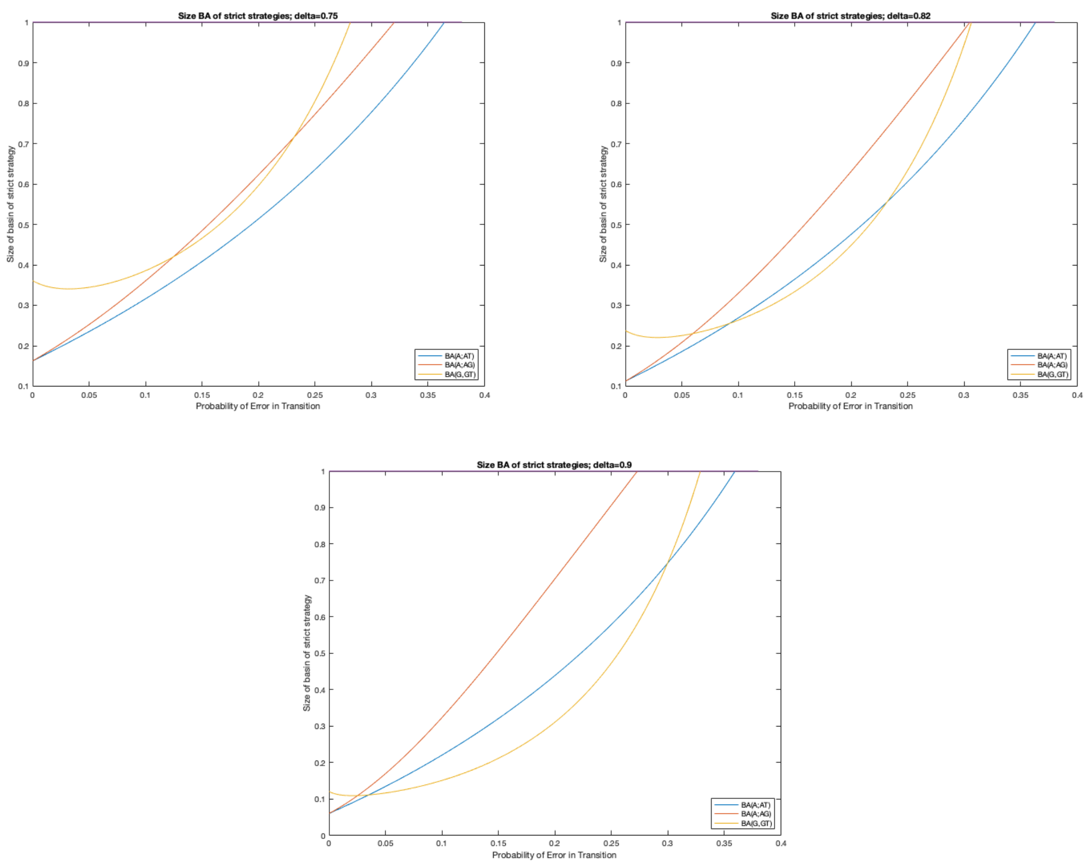


Figure 5: **Proportional Imitation dynamics with no error.** All points on the segment joining s_{AG} and s_{AT} (denoted by U) are unstable steady states. All points in the segment joining G and T are stable steady states. All lines joining A with points on the latter segment are invariant. Only few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

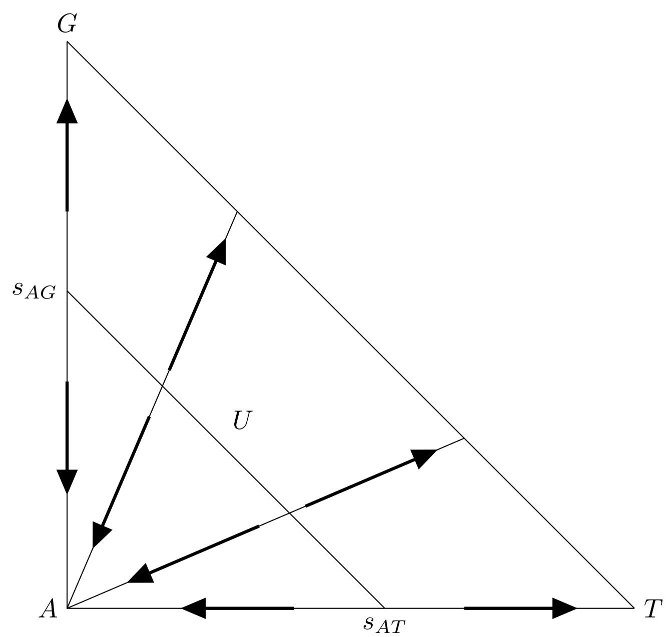


Figure 6: **Flow after the first transition: Low δ.** Here $\epsilon > \bar{\epsilon}$, but the two values are close. Also $\delta < \hat{\delta}$, hence the steady state s_{GT} disappears first. There are only two sub-regions of the interior of the projection of the simplex, with attractors A and G.

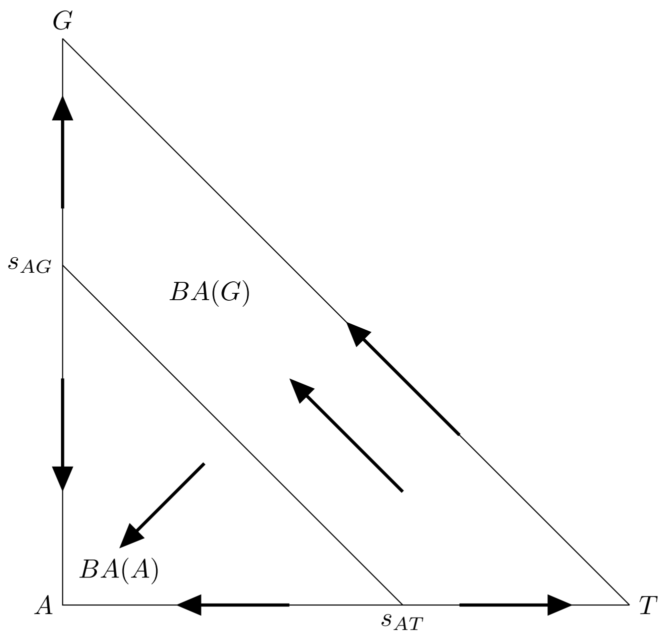


Figure 7: **Frequency of Strategies over Time.**

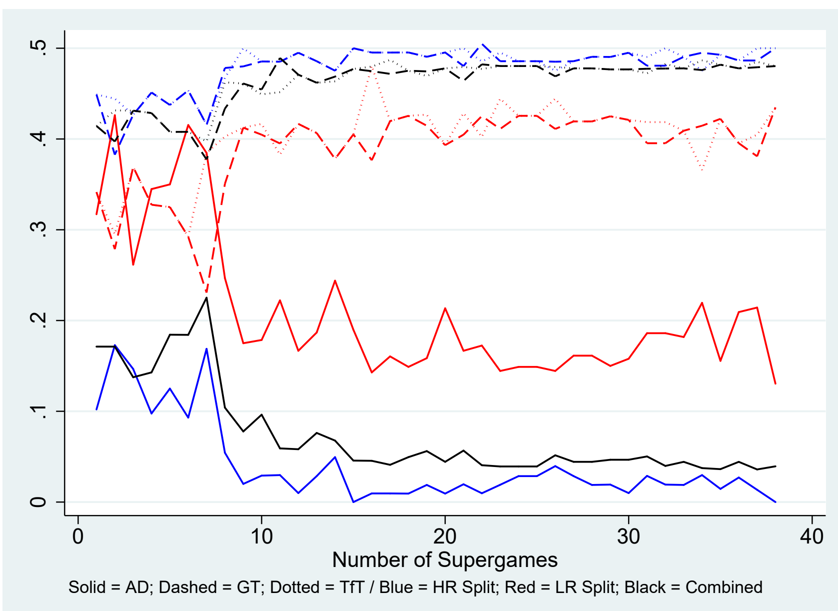


Table 1: **Prisoner's Dilemma.** C: Cooperate, D: Defect.

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

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 the 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 supergames** 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 effects 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 effects (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: **Outcomes at period $t-1$ as determinants of cooperative choice 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 b/p	High IQ Split b/p	Low IQ Combined b/p	High IQ Combined 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 b/p	High IQ Split b/p	Low IQ Combined b/p	High IQ Combined 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 6: **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 for 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 All b/se	Low IQ All 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 7: **Individual strategies and errors in the different treatments.**

Treatment	**IQ Split**		**Combined**		
IQ Session/Group	High	Low	All	High	Low
Strategy					
A	0.073	0.331	0.130	0.100	0.160
G	0.461	0.325	0.433	0.448	0.419
T	0.465	0.343	0.437	0.452	0.422
Ratio G/T	0.991	0.947	0.992	0.991	0.993
Ratio A/T	0.158	0.965	0.297	0.221	0.379
Ratio A/G	0.159	1.019	0.299	0.223	0.381
Error in transition (ϵ)	0.005	0.011	0.007	0.007	0.008
Error in action rates (η)	0.010	0.020	0.010	0.009	0.012

Table 8: **Individual strategies and errors in the different treatments.**

Supergames	**First 10**			**Last 10**		
Treatment	IQ Split		Combined	IQ Split		Combined
IQ Session/Group	High	Low	All	High	Low	All
Strategy						
A	0.170	0.445	0.248	0.038	0.297	0.080
G	0.409	0.261	0.370	0.479	0.350	0.458
T	0.421	0.294	0.382	0.483	0.353	0.461
Ratio A/T	0.405	1.512	0.648	0.078	0.841	0.174
Ratio A/G	0.416	1.702	0.669	0.078	0.850	0.175
Ratio G/T	0.972	0.889	0.968	0.991	0.990	0.993
Error in transition rates (ϵ)	0.008	0.015	0.013	0.005	0.004	0.005
Error in action rates (η)	0.018	0.028	0.021	0.008	0.010	0.006

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.

Bordalo, P., Gennaioli, N. and Shleifer, A. (2017). Memory, attention, and choice, *Technical report*, National Bureau of Economic Research.

Camerer, C. F., Ho, T.-H. and Chong, J.-K. (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games, *Journal of Economic theory* **104**(1): 137–188.

Cason, T. N., Lau, S.-H. P. and Mui, V.-L. (2013). Learning, teaching, and turn taking in the repeated assignment game, *Economic Theory* **54**(2): 335–357.

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.

Dal Bó, P. and Fréchette, G. R. (2019). Strategy choice in the infinitely repeated prisoner's dilemma, *American Economic Review* **109**(11): 3929–52.

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.

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories, *Handbook of the economics of giving, altruism and reciprocity* **1**: 615–691.

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.

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.

Jones, G. (2008). Are smarter groups more cooperative? evidence from prisoner's dilemma experiments, 1959–2003, *Journal of Economic Behavior & Organization* **68**(3): 489–497.

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.

Proto, E., Rustichini, A. and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions, *Journal of Political Economy* **127**(3): 1351–1390.

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.

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.

Appendices

A Details on Design and Implementation

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.

[18] See note 2.

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.

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 0.003 Euro.

4. A de-briefing questionnaire was administered.

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

B Session Dates, Sizes and Characteristics

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.

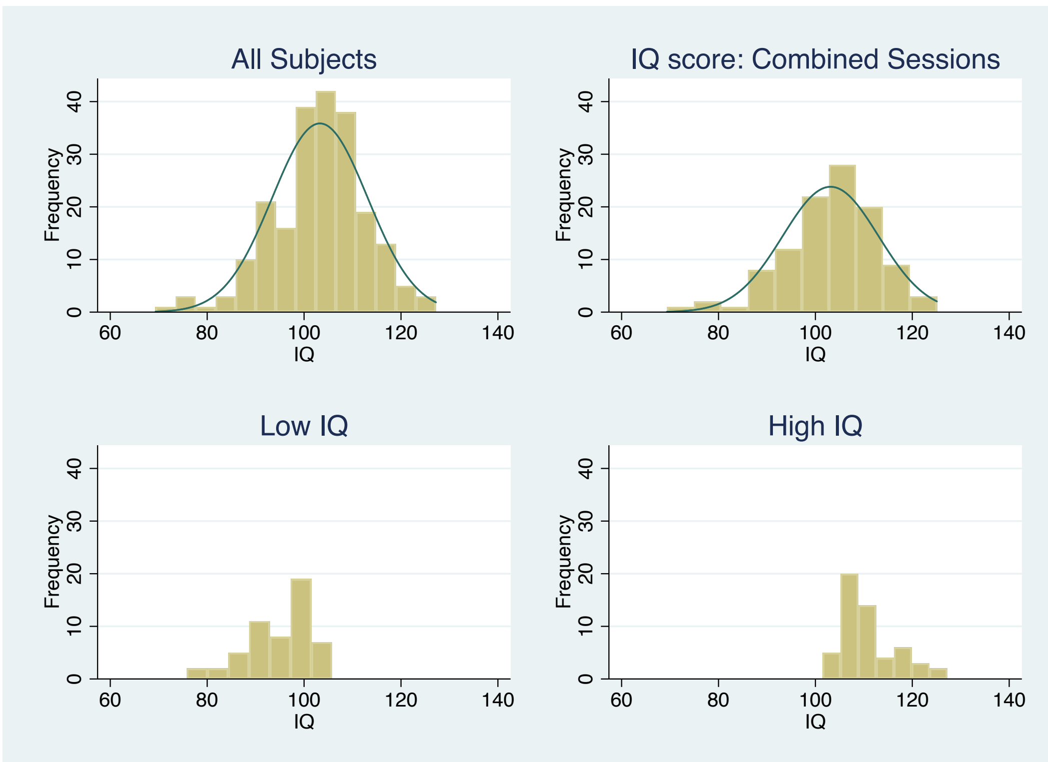


Table A.6: **Countries of Origin of Participants**

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

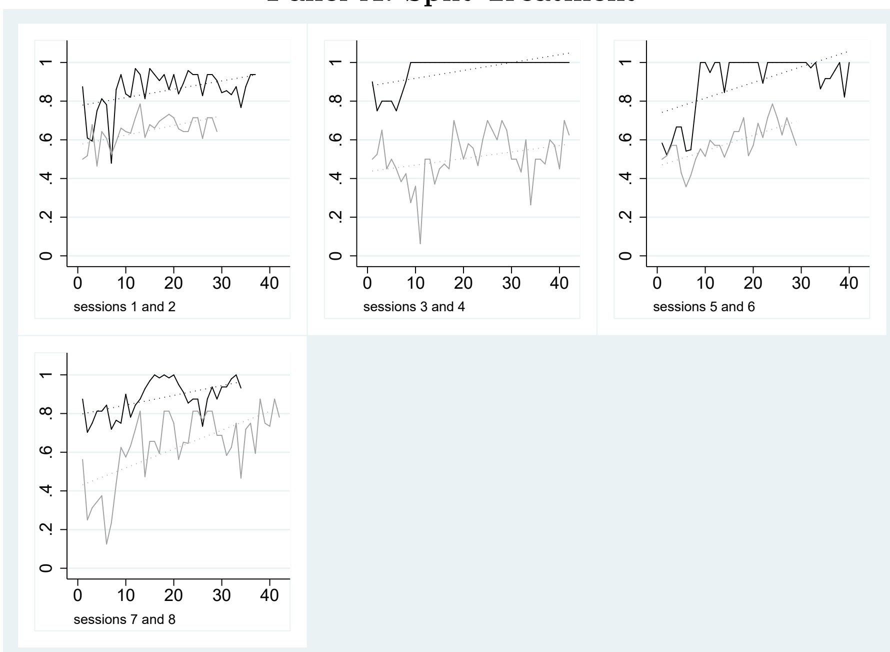
Table A.14: All participants: Correlations Table (p−*values* in brackets)

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)	

C Supplementary Data Analysis

Figure A.2: **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

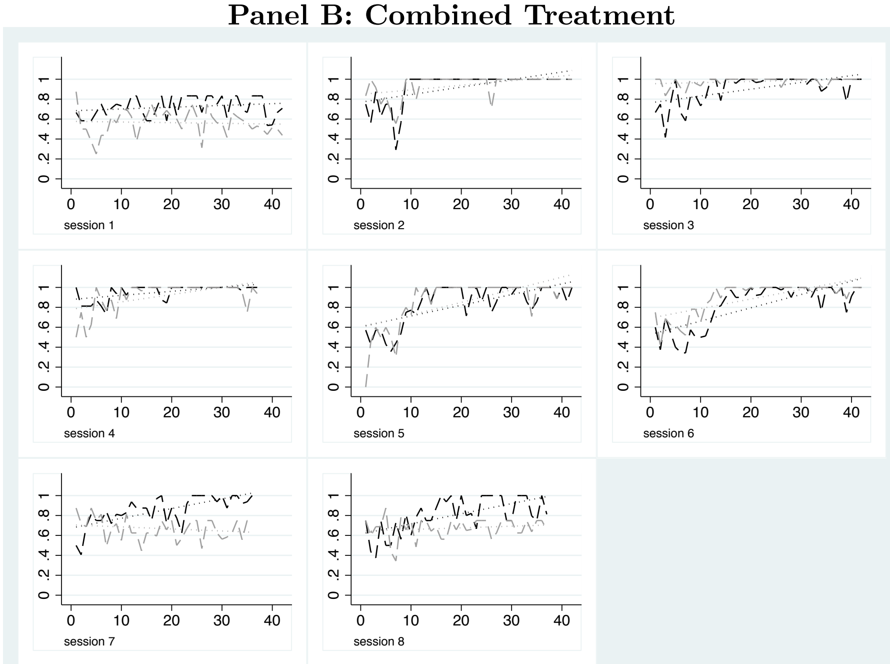
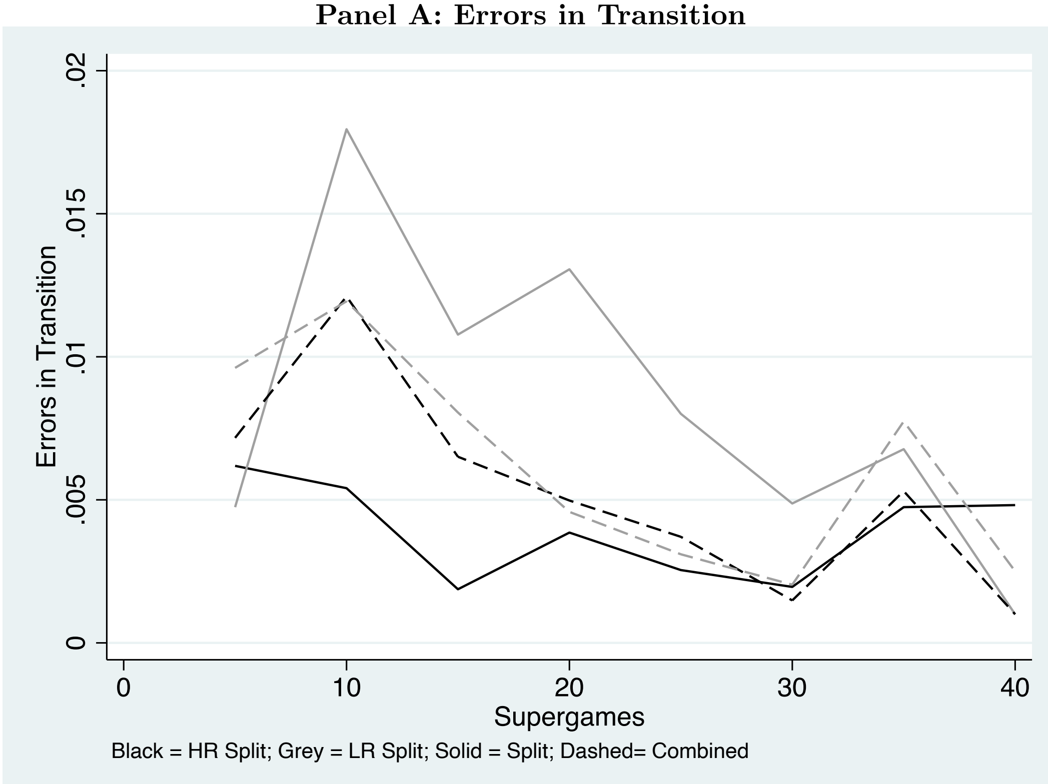


Figure A.3: **Errors in transition and action,** Error rates across different treatments and groups. The average is computed over observations in successive blocks of five supergames. The grey solid line represent the low-IQ split treatment, the black solid lines the high-IQ split treatment, the dashed lines the combined treatment: grey is the low-IQ group, black is the high-IQ group.

Panel A: Errors in Transition



Panel B: Errors in Action

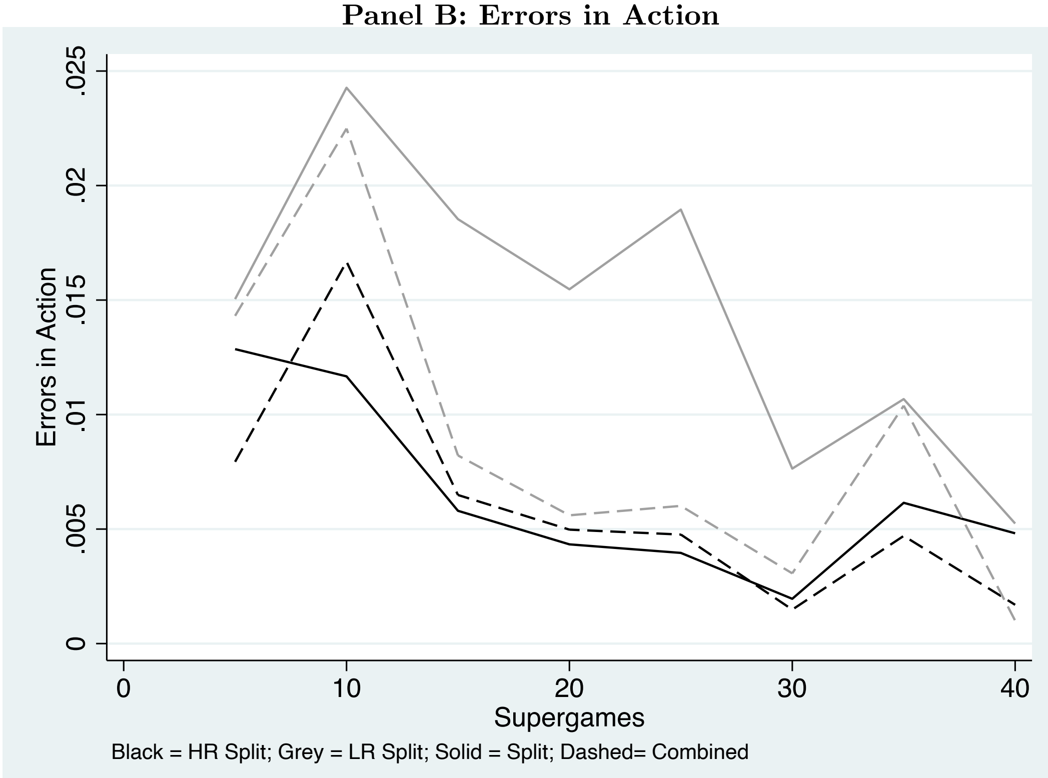


Figure A.4: **Frequency of Transition Errors.** Top panel: all supergames; bottom panel: early supergames (first half).

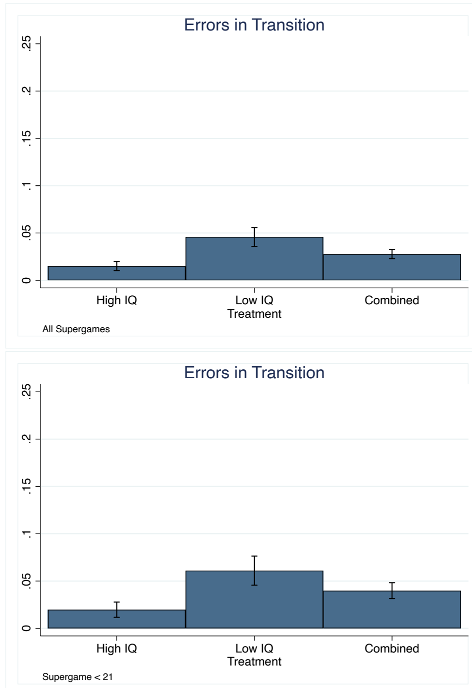


Figure A.5: **Frequency of Action Errors.** Top panel: all supergames; bottom panel: early supergames (first half).

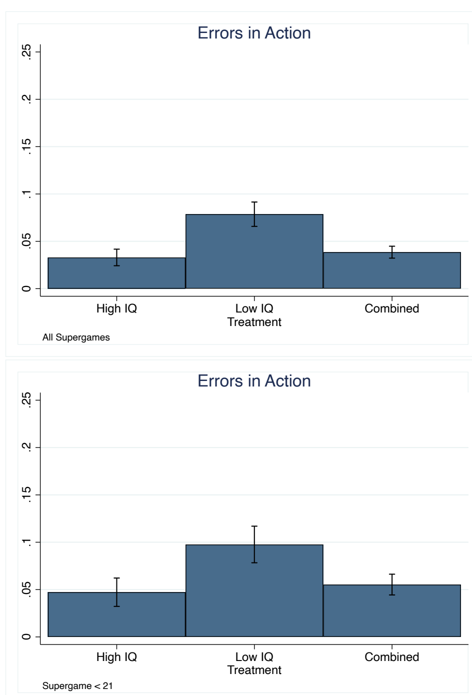


Figure A.6: **Frequency of Transition Errors: Histogram.** Early supergames, common x and y-scale

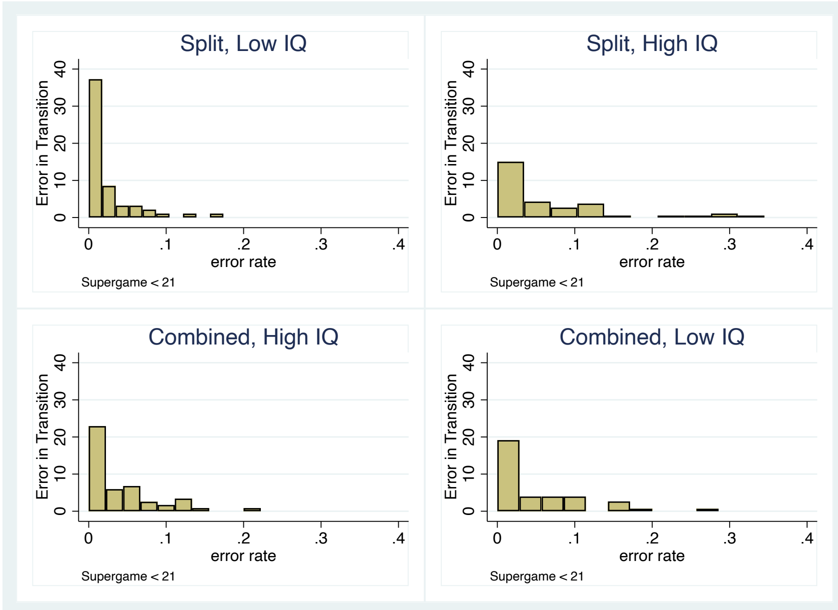
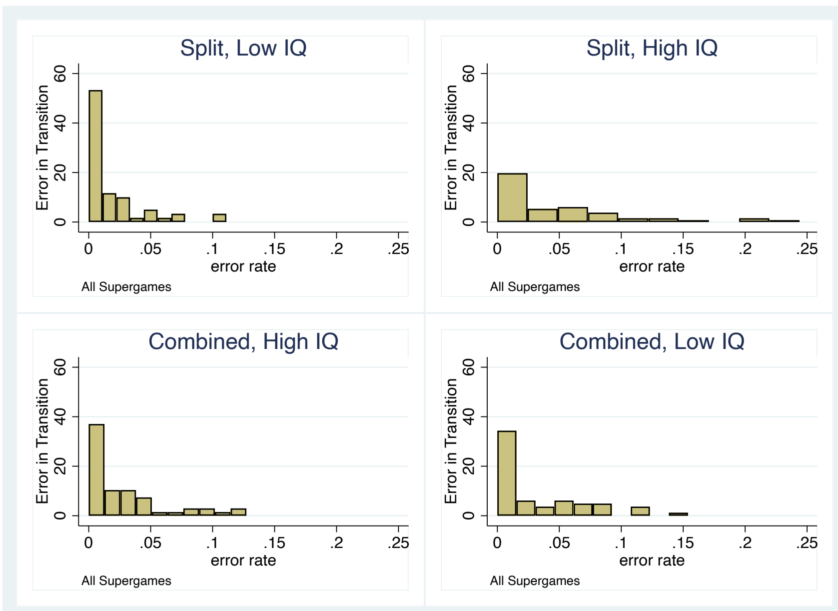


Figure A.7: **Frequency of Transition Errors: Histogram.** All supergames, common x and y-scale.



D Payoff Tables with Transition Errors

We report the value of the matrix V for the values of δ relative to Figure (4), top and bottom panel, for various values of ϵ. The entries illustrate the transition along different types of equilibria and attractors.

In Table (A.15), 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 (6) 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.15: V **matrix,** $\delta = 0.75$**,** $\epsilon = 0.25, 0.30, 0.35, 0.40$**.**

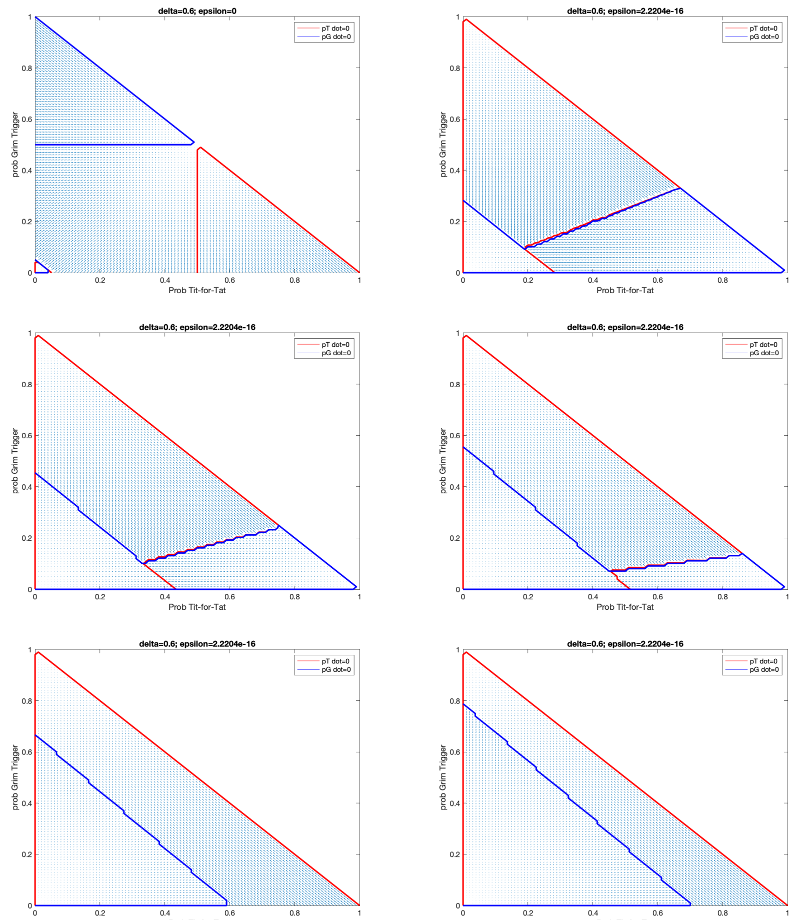
	A	G	T
A	25.0000	35.9375	35.9375
G	19.3125	37.6125	39.1168
T	19.3125	37.2846	39.1962
	A	G	T
A	25.0000	36.8750	36.8750
G	18.8250	37.3221	38.6771
T	18.8250	37.0298	38.6384
	A	G	T
A	25.0000	37.8125	37.8125
G	18.3375	37.1935	38.3034
T	18.3375	36.9510	38.1969
	A	G	T
A	25.0000	38.7500	38.7500
G	17.8500	37.1718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

In Table (A.16), 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.

Table A.16: V **matrix,** $\delta = 0.9$**,** $\epsilon = 0.25, 0.30, 0.35, 0.40$**.**

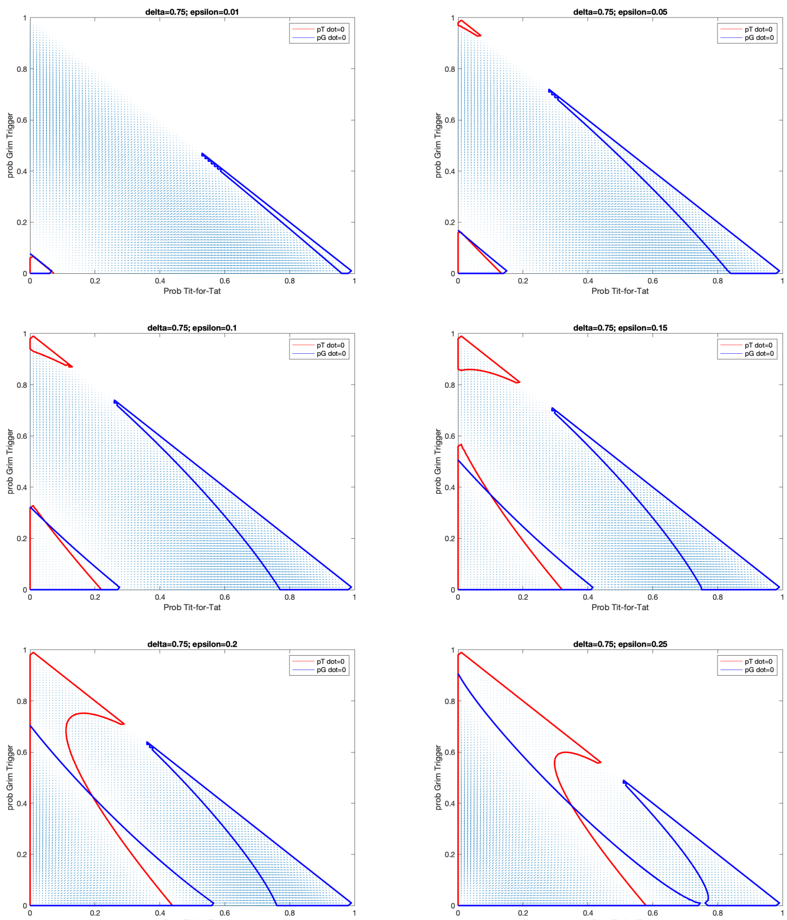
	A	G	T
A	25.0000	33.1250	33.1250
G	20.7750	33.5591	35.7744
T	20.7750	33.1806	36.1957
	A	G	T
A	25.0000	34.2500	34.2500
G	20.1900	33.7727	35.7489
T	20.1900	33.4154	35.8681
	A	G	T
A	25.0000	35.3750	35.3750
G	19.6050	34.0813	35.6889
T	19.6050	33.7695	35.6246
	A	G	T
A	25.0000	36.5000	36.5000
G	19.0200	34.4359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

Figure A.8: **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 0 to 0.25. Payoffs are as in our experimental design; discount factor is $\delta = 0.6$.



E Vector Field with Replicator Dynamics

Figure A.9: **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.



F 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

$$A_\eta(\omega)(\omega', a) \equiv \delta_\omega(\omega')Pr(a;\omega,\eta) \tag{A-1}$$

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-2). 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:

$$Q_{\epsilon,\eta}(\omega';\omega) \equiv \sum_{a\in A} A_\eta(\omega)(\omega, a)T_\epsilon(\omega, a)(\omega') \tag{A-2}$$

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

$$V_{\epsilon,\eta} = (1-\delta)u_\eta + \delta Q_{\epsilon,\eta}V_{\epsilon,\eta} \tag{A-3}$$

F.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 F.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 into invariant sets holds in the current case as well. Similarly, with easy computations, one gets:

Lemma F.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 F.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, also in the case of the two errors A survives for all values of the parameters.

F.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.10) and (A.11) report the sizes of the basin as function of the two error rates. These are the three dimensional versions of the figure (4).

One can consider also the exact analogue of figure (4), 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.10: **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).

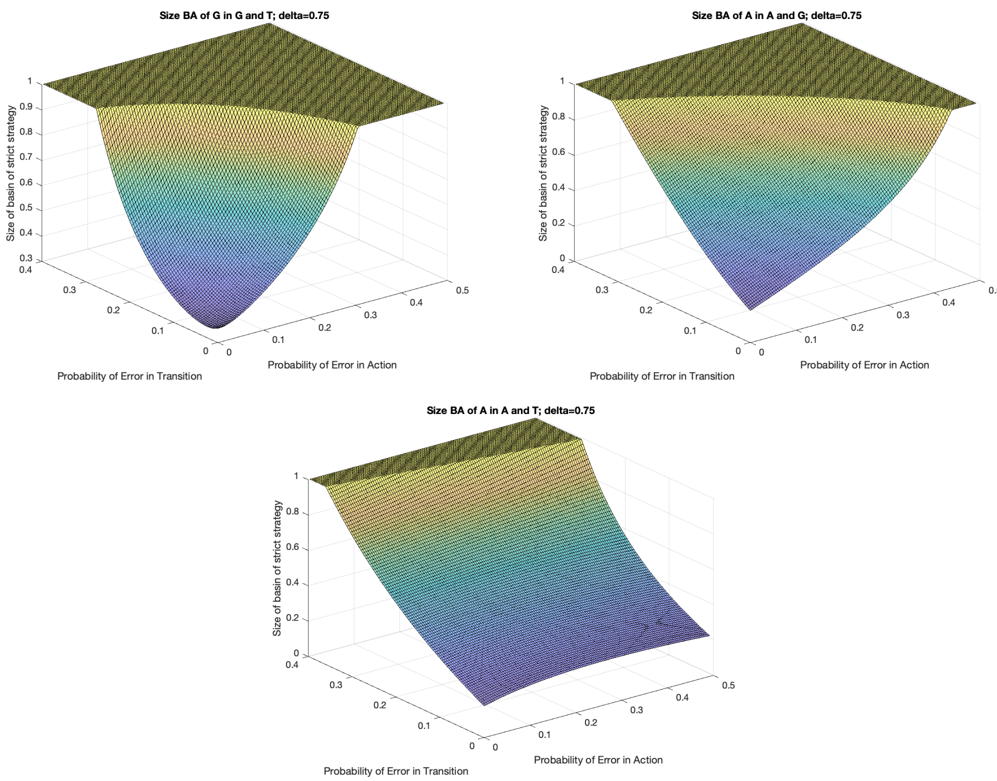
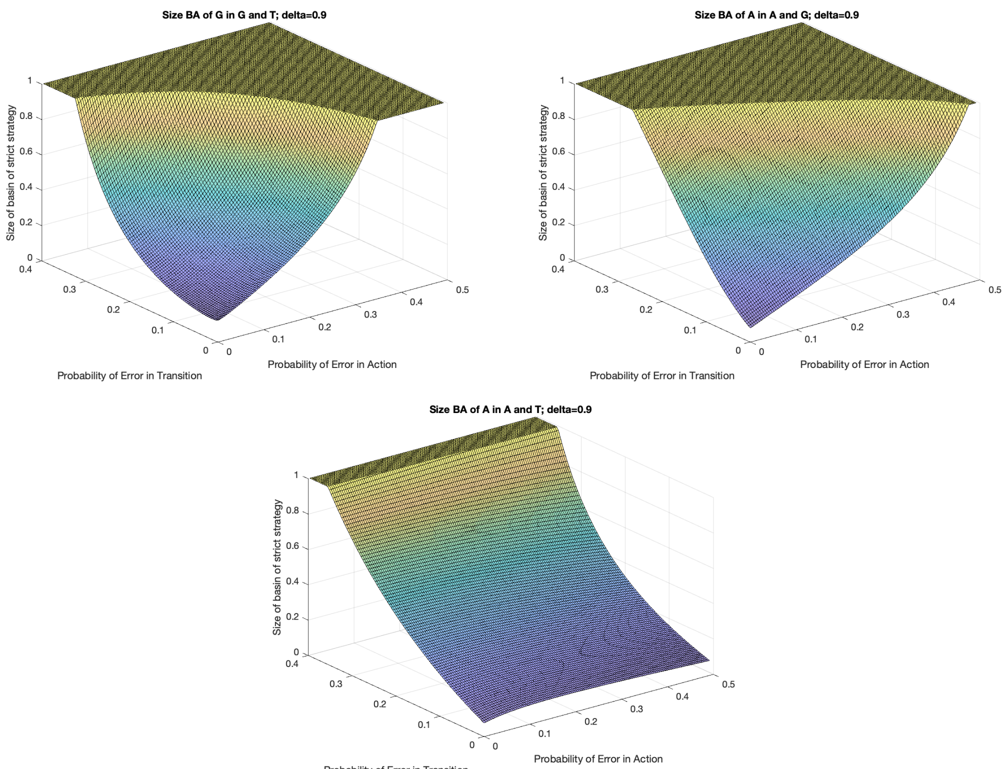


Figure A.11: **Size of basin of attraction of strict strategies with two types of error.** Probability of error in transition and action choice as displayed; $\delta = 0.9$. Top to bottom panel: values of the basin of attraction in the indicated subsets of strategies (basin of G in GT, A in AG, A in AT, respectively).



G Learning Model

G.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 at every point in time a strategy profile he adopts, call it $s^i(t)$. Such assignment 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.

G.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:

$$\forall s \in S : \frac{d\mu(s,t)}{dt} \in \lambda \left(BR(\mu(\cdot,t))(s) - \mu(s,t) \right). \tag{A-4}$$

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

$$\mu(s,t+dt) \in (1 - \lambda dt)\mu(s,t) + \lambda dt BR(\mu(\cdot,t))(s) \tag{A-5}$$

which in the limit gives the differential inclusion (A-4).

We now apply the population model to a learning model.

G.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:

$$D(\mu,\alpha^i) = \frac{1}{B(\alpha^i)} \mu(A)^{\alpha^i_A - 1} \mu(G)^{\alpha^i_G - 1} \mu(T)^{\alpha^i_T - 1} \tag{A-6}$$

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.

G.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:

$$\pi \in \Delta(\mathbb{Z}^3) \tag{A-7}$$

G.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:

$$BR(\alpha) = argmax_{\sigma \in \Delta(S)} E_{D(\cdot,\alpha)} u(\sigma, \cdot). \tag{A-8}$$

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

$$E_{D(\cdot,\alpha)}\mu(s) = \frac{\alpha_s}{\sum_{r \in S} \alpha_r}.$$

$$\equiv M(s;\alpha) \tag{A-9}$$

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

$$BR(\alpha) = argmax_{\sigma \in \Delta(S)} \sum_{s \in S} M(s;\alpha) u(\sigma, s). \tag{A-10}$$

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. [19]

[19] 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:

$$\phi(s;\pi) \equiv \pi(\{\alpha : BR(\alpha) = \{s\}) + \sum_{r \neq s} \frac{1}{2} \pi(\{\alpha : BR(\alpha) = \{s, r\}) + \tag{A-11}$$

$$\frac{1}{3} \pi(\{\alpha : BR(\alpha) = \Delta(S))$$

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

$$\pi(\alpha, k+1) = S\pi(\cdot, k), k = 0, 1, \ldots. \tag{A-12}$$

where S defined in (A-14).

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:

$$\alpha_s^{i,\prime} = \alpha_s^i + \delta_s(b^i) \tag{A-13}$$

where δ_s is the indicator function. [20] Define:

$$(S\pi)(\alpha) \equiv \sum_{s \in S} \pi(\alpha - 1_s)\phi(s, \pi) \tag{A-14}$$

Recall our discussion before (A-7), so the definition (A-14) is meaningful even when, in $\alpha_s = 0$. Equation (A-12) follows. □

G.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 (G.1.1), with the distribution on beliefs π replacing the distribution on strategies μ:

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

$$\frac{d\pi(\alpha, t)}{dt} = \lambda \left(\sum_{s \in S} \pi(\alpha - 1_s)\phi(s, \pi(\cdot, t)) - \pi(\alpha, t) \right) \tag{A-15}$$

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-13). 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

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

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-15) 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.

G.3 Estimation of beliefs' updating under best response

We now estimate the evolutions of the beliefs under best response dynamics described above for each strategy, s, at the beginning of each supergame t, $\alpha^i_{s,t}$. For simplicity we limit ourselves to the case of no errors in the implementation of strategies (transition or action), where G and T have the same expected utility and for this reason we refer to this strategy as *sophisticated cooperation* or SC. Therefore, we assume that subjects in the first repeated game hold beliefs that other players either use A or a cooperative strategy that we already defined SC. Let the probability of player i in supergame s to play A be $\alpha^i_{A,t}/(\alpha^i_{A,t} + \alpha^i_{SC,t})$. In the first supergame, $t = 1$, subjects have beliefs charaterized by $\alpha^i_{A,1}$ and $\alpha^i_{SC,1}$, from the second supergame onward, $t > 1$. Following Dal Bó and Fréchette (2011), we assume that they update their beliefs as follows:

$$\alpha^i_{k,t+1} = \theta_i \alpha^i_{k,t} + 1(a^j_k), \tag{A-16}$$

where k is the action (A 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

$$U^i_{a,t} = \frac{\alpha^i_{A,t}}{\alpha^i_{A,t} + \alpha^i_{SC,t}} u_a(a^j_A) + \frac{\alpha^i_{SC,t}}{\alpha^i_{A,t} + \alpha^i_{SC,t}} u_a(a^j_{SC}) + \lambda^i_s \epsilon^s_{a,t} \tag{A-17}$$

where $u_a(a^j_k)$ 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.12. We now analyse the two parameters we are interested: θ_i, measuring the inverse of the speed by which subjects update their beliefs and λ^i_s, measuring the inverse of the capacity of best responding given the beliefs.[21]

In table A.19, 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.13 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.18 , 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.13 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.18 confirm that the differences among the low IQ in the combined and in the split treatments are statistically significant. We can summarise this 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.

[21] We omit the details on how the model is estimated, they can be found in in the online appendix of Dal Bó and Fréchette (2011) at page 6-8.

Figure A.12: **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.

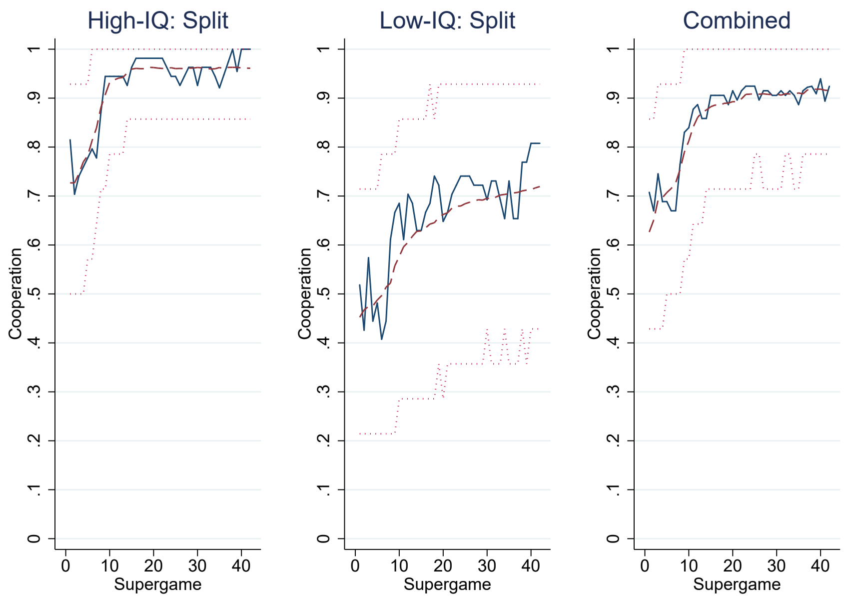


Figure A.13: **Distribution of the beliefs' updating speed within the different groups and treatments**. Distribution of the parameter θ_i as defined in equation A-16, where 1 correspond to slowest speed (fictitious play) and 0 to the fasted speed (Cournot dynamics)

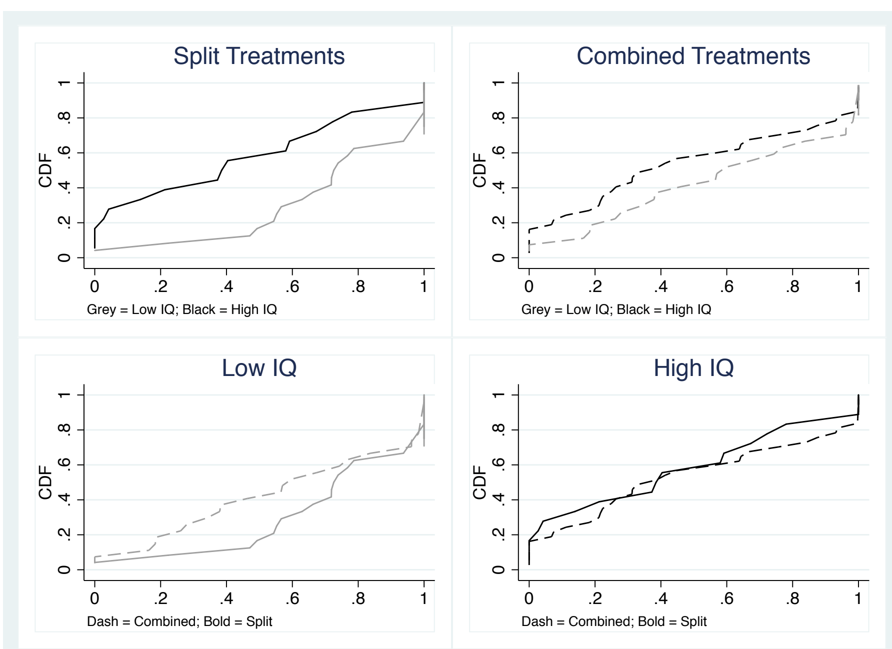


Table A.17: IQ and Simulated Parameters

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.18: **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-16, 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.19: **Correlation between IQ, beliefs updating and capacity of best responding to own beliefs** Correlations between IQ, updating speed, θ_i, capacity of best responding to beliefs in supergame s, λ_s. $p-values$ 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)	

H Estimation of the error rates

Let s denote the strategy in a given set of strategies (to be determined later), ϵ the probability of an error in action choice and η the probability of an errors in transition among states of the automaton.

Take a history of the supergame of length T, $h \equiv (a_1^1, a_1^2, \ldots, a_T^1, a_T^2)$. We are interested in determining the parameter $(s, \epsilon, \eta) \in S \times [0, 0.5]^2$ that maximizes the probability of observing the choices of subject 1 given the history of the others' actions. The restriction to errors less than chance is natural. The model we have described in the earlier section provides a statistical model which we now describe.

H.1 Recursive Computation of Errors

There is a simple recursive algorithm to compute the set of all possible error histories that explain the observed choices for a given strategy. The algorithm goes through the sequence of observations, in the order of occurrence, and produces at every step a set of partial records which are triples $(n_1, n_2, \omega) \in \mathbb{N}^2 \times \Omega$, where n_1 is the number of errors in action, n_2 is the number of errors in transition, ω is the internal states of the automaton. So the interpretation of the triple (n_1, n_2, ω) is:

> There exists a sequence of errors with a total of n_1 errors in action and n_2 errors in transition explains the observations up to the current observation; according to this sequence, the current internal state of the automaton representing the strategy is ω.

We first illustrate this algorithm in the case of the observations (A-18), considering $s = G$, with internal states $\{c, d\}$.

$$(C^1C^2; D^1C^2; C^1C^2) \tag{A-18}$$

1. Consider the first period. For each record we have we do the following:
   (a) **(Prediction of action and comparison of observation and prediction.)** The initial record is $(0, 0, c)$. This predicts an action C^1, which is what we observe, so the record is unchanged
   (b) **(Updating of the state given the history)** Since the action of the opponent is C^2, the state at the beginning of the second period should be c according to this partial record (the only one we have), and the action produced C^1. At this final step for period 1 there is only one partial record, $(0, 0, c)$
   (c) **(Go to next.)** Go to the next observation.
2. Consider the second period. For each record we have we do the following:

(a) **(Prediction of action and comparison of observation and prediction.)** The only record we have predicts an action C^1. If we did observe C^1 we would proceed to the step of transiting to the new state After observing D^1 instead, the algorithm produces two records: $(1, 0, c)$ (an error in action and the internal state unchanged, as it should because both players cooperated in the first period), and $(0, 1, d)$ (an error in transition and the state changed to d). Note that c and d in the two partial records respectively refer to what the state was *at the end of period 1*. In the case of record $(0, 1, d)$ for example we are considering the following explanation: upon observing C^1C^2 at the end of the first period the state should have transited to (that is, remained at) c; instead it transited to d.

(b) **(Updating of the state given the history)** Now we perform on each partial record the updating of the state: given that the observed action profile is D^1C^2 the two records $(1, 0, c)$ and $(0, 1, d)$ should change to $(1, 0, d)$ and $(0, 1, d)$ respectively.

(c) **(Go to next.)** Go to the next observation.

We then proceed with the next observation, and execute the same procedure for each possible partial record. Each of them may produce a set of possible "offsprings", and this occurs if and only if an error (according to the prediction of the partial record) is observed. If no error is observed, the state should be updated at the strategy requires, depending on the action profile. Note that whether an action is an error or not depends on the partial record.

1. **(Prediction of action and comparison of observation and prediction.)**

   (a) The record $(1, 0, d)$ produces two new records, $(1, 1, c)$ and $(2, 0, d)$.

   (b) The record $(0, 1, d)$ produces two new records, $(1, 1, d)$ and $(0, 2, c)$.

2. **(Updating of the state given the history)**: This step is now irrelevant, because we have reached the final period. If this was not the final period, we would have for the next period:

   (a) The two records $(1, 1, c)$ and $(2, 0, d)$ should produce $(1, 1, c)$ and $(2, 0, d)$.

   (b) The two records $(1, 1, d)$ and $(0, 2, c)$ should produce $(1, 1, d)$ and $(0, 2, c)$.

   and we would continue the process for these

H.2 General Algorithm

In summary the algorithm for player 1 is:

1. We have a data set $(a_1b_1, a_2b_2, \ldots, a_tb_t, \ldots, a_Tb_T)$ (denoting a the action of the first player, b of the second);

2. We consider each possible automaton m in some candidate set M; here for simplicity of exposition we consider the case of automata with only two states;

3. For each of these strategies, for the period t we have a set of partial records $(n_1^k, n_2^k, \omega^k) : k = 1, \ldots, K_t$;

4. For each k record and each period t:

   (a) **(Prediction-comparison)**: Given the state ω^k in the k^{th} record and the strategy m, predict the action $\tilde{a}_{t+1}$, and check whether $a_{t+1} = \tilde{a}_{t+1}$:

      i. if $a_{t+1} = \tilde{a}_{t+1}$, keep the records and proceed to the updating step below;

      ii. if $a_{t+1} \neq \tilde{a}_{t+1}$, produce two new records:

      $$(n_1^k + 1, n_2^k, \omega^k)$$

      corresponding to an error in action, and

      $$(n_1^k, n_2^k + 1, (\omega^k)')$$

      corresponding to an error in transition; here $(\omega^k)'$ is the opposite than state ω^k. Delete the record (n_1^k, n_2^k, ω^k). Recall that ω^k and $(\omega^k)'$ and should be interpreted as our guess at what the state at the beginning of period $t+1$ must have been.

   (b) **(Updating)**: For each new record, update the state according to the observed actions $(a_{t+1}b_{t+1})$ and the transition rule of m.

   (c) **(Go to next.)** Go to the next record and then to the next period.

H.2.1 Remark

The algorithm applied to the T strategy produces records that have the sum of errors in action and transition equal for all records. This fact follows because the transition in the automaton only depends on the action of the other. By induction consider a record (n_1^k, n_2^k, ω^k). If the observed action is different from the predicted one, then at the next step the algorithm produces two new records $(n_1^k + 1, n_2^k, \omega^k)$ and $(n_1^k, n_2^k + 1, \omega^{k,\prime})$. The difference in state of the automaton is erased by the subsequent updating that only depends on the action of the other player, hence it is the same for the two records we are considering. Thus the two new records have the same sum of errors.

H.3 Simultaneous Errors

In the estimate of the error rates ϵ and η we want to assume that the two errors are independent; so the possibility that two errors occur in the same period is small, $\epsilon\eta$, but positive. We note in this subsection that if we are interested in the maximum likelihood estimation, we can ignore this possibility.

The algorithm we described only considers the possibility of a single error in each period. We need to examine the possibility that an algorithm which allows two errors in the same period. Although it produces a strictly larger number of errors in the current period for at least on type of error, this algorithm might produce a better report in the final stage. The next proposition shows that this is not the case.

Proposition H.1. *For any history and any final report obtained allowing two errors (one in action and one in transition) there is a report that can be obtained allowing only one error in each stage that has a weakly smaller number for each action and transition.*

Proof. Fix the strategy, and its automaton, that we are considering; and fix a history of the supergame of length T, $h \equiv (a_1^1, a_1^2, \ldots, a_T^1, a_T^2)$. Let

$$(n_t^1, n_t^2, s_t) \equiv r_t$$

indicate a report at t, after updating the state of the automaton (that is after the *Updating* step of section H.2 is concluded). We define a *continuation set* at report $r = (n^1, n^2, s_t)$ and partial history

$$h_{t+1} \equiv (a_{t+1}^1, a_{t+1}^2, \ldots, a_T^1, a_T^2),$$

denoted $C(r, h_{t+1})$, as the set of final reports that can be generated with that history and initial report by the algorithm which allows a single error in every period. We will refer to the two algorithms as single error and double error for short. It is clear that the continuation set only depends on the state in the report, not on the recorded number of errors. So if we prove that the single error report generates in every t the same set of states with a weakly smaller (for each type of error) number of errors, then our claim will follow. We indicate by T the transition function of the strategy we are considering.

We consider two cases:

1. $a_{t+1}^1 \neq s_t$. In this case the single error algorithm produces two continuations reports, $(n^1, n^2, T(s_t, a_{t+1}^1, a_{t+1}^2))$ and $(n^1, n^2, T(s_t', a_{t+1}^1, a_{t+1}^2))$, where the prime indicates the other state, that is $s_t \neq s_t'$. So before the updating both states are reached; so the set of states that are reached after the updating in the single error algorithm contains the set of states that can be reached with the double error algorithm, with a weakly smaller error vector.

2. $a_{t+1}^1 = s_t$. In this case the single error algorithm produces a single report:

$$(n^1, n^2, T(s_t, a_{t+1}^1, a_{t+1}^2)) \equiv r_{t+1},$$

whereas the double error algorithm produces a report

$$(n^1 + 1, n^2 + 1, T(s_t', a_{t+1}^1, a_{t+1}^2)) \equiv r_{t+1}',$$

which might not be produced by the single error algorithm, at the cost of one additional error of each type. There is however no error to explain at t. Suppose that an error arises at any later period according to the report following r_{t+1} and the intervening history, and instead no error arises according to the report that follows from r'_{t+1} and the intervening history. That error can be explained in the single error algorithm adjusting the the report following r_{t+1} (if needed, because the state is different from the one producing the observed action) with a single error in transition.

□

H.4 Estimate of Error rates

We can now proceed to the estimation of the error rates. This step is standard maximum likelihood estimation over the set $(s, \epsilon, \eta) \in S \times [0, 0.5]^2$ of strategies and error rates. The final reports (if we ignore the state of the automaton, which is irrelevant in the final step) have the form (n^1, n^2), the number of each type of error, for a subgame of length T and for each strategy; so we may write the total allocation of periods to different types of errors (possibly none) (n^Z, n^A, n^T, n^{AT}), for no-error, error in action, errors in transition, double error. Since there are no double errors, and the total sum is T, the log-likelihood of the (ϵ, η) pair at $(T - n^A - n^T, n^A, n^T, 0)$, assuming independence of the errors, is

$$(T - n^T)log(1 - \epsilon) + n^T log(\epsilon) + (T - n^A)log(1 - \eta) + n^A log(\eta)$$

so the ML estimated parameters are the relative frequencies:

$$\hat{\epsilon} = \frac{n^T}{T}, \hat{\eta} = \frac{n^A}{T}. \tag{A-19}$$

We then choose the maximum over the strategies. If the maximum likelihood procedure attributes the same probability to two different strategies, we assume that both can be chosen with equal probability.[22]

[22] This is irrelevant in the computation of the individual error rates since when two strategy have the same likelihood in a supergame the two records will have the same error rates by construction.

Table A.20: **Determinants of error rates across individuals.** Dependent variables are the error rates standardized. High IQ split is the baseline treatment indicator in columns 1 and 3. Tit for tat is the baseline strategy. GLS estimator with random-effect. Errors are clustered at the individual levels; * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	Transition Error rates b/se	Transition Error rates b/se	Action Error rates b/se	Action Error rates b/se
Low IQ split	0.1987***		0.0655	
	(0.0564)		(0.0778)	
Combined	0.0795**		–0.0252	
	(0.0325)		(0.0510)	
IQ		–0.0044**		–0.0046*
		(0.0019)		(0.0025)
Always Defect			0.0668	0.0661
			(0.1878)	(0.1857)
Grim Trigger	0.2908*	0.2644	–0.9384***	–0.9409***
	(0.1655)	(0.1647)	(0.3475)	(0.3478)
Supergame	–0.0075***	–0.0075***	–0.0051***	–0.0052***
	(0.0012)	(0.0012)	(0.0013)	(0.0013)
N	7492	7492	7492	7492

Table A.21: **Individual strategies in the different treatments using Dal Bo and Frechette ML estimation**. The details on how this table is estimated are in the online appendix of Dal Bó and Fréchette (2011).

Supergames	**First 10**			**Last 10**		
Treatment	IQ Split		Combined	IQ Split		Combined
IQ Session/Group	High	Low	All	High	Low	All
Strategy						
A	0.100	0.365	0.168	0.037	0.228	0.076
G	0.423	0.310	0.404	0.295	0.340	0.444
T	0.476	0.324	0.427	0.669	0.438	0.480
Proportion A/T	0.210	1.127	0.393	0.055	0.521	0.157
Proportion A/G	0.236	1.177	0.416	0.126	0.671	0.170
Proportion G/T	0.889	0.957	0.946	0.441	0.776	0.925