Intelligence, Errors and Cooperation in Repeated Interactions

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

March 22, 2021

Abstract

We study how strategic interaction and cooperation are affected by the heterogeneity of cognitive skills of groups of players, over consecutive plays of repeated games with randomly matched opponents using Prisoner's Dilemma as stage game. We observe overall higher cooperation rates and average final payoffs in integrated treatment groups – where subjects of different IQ levels interact together– than in separated treatment groups. Lower IQ subjects are better off and higher IQ subjects are worse off in integrated groups than in separated groups. Higher IQ subjects adopt harsher strategies when they are pooled with lower IQ subjects than when they play separately. We demonstrate that this outcome should be expected in learning and evolutionary models where higher intelligence subjects exhibit lower frequency of errors in the implementation of strategies. Estimations of errors and strategies in our experimental data are consistent with the model's assumptions and predictions.

JEL classification: C73, C91, C92, D83
Keywords: Evolutionary Games, Learning, Repeated Prisoners Dilemma, Cooperation, Intelligence, Strategy Errors

---

[*]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, Dimitris Christelis, Drew Fudenberg, David Gill, Marco Lambrecht, Fabio Maccheroni, Salvatore Nunnari, Joerg Oechssler, Emel Ozbay, Erkut Ozbay, Antonio Penta, Andy Schotter. Larbi Alaoui and Guillaume Fréchette have provided penetrating comments that we tried to address to the best of our ability. A discussion of the model with David Levine was extremely 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 and many other audiences. Responsibility for errors and confusion is ours. We finally thank four anonymous referees and the editor for comments that have considerably helped to improve the analysis and presentation. The University of Heidelberg provided funding for this research. AR thanks the National Science Foundation, grant NSF 1728056.

1 Introduction

Intelligence is an important characteristic affecting single agent and strategic behaviour (e.g. Jones, 2008; Alaoui and Penta, 2016; Gill and Prowse, 2016; Alaoui and Penta, 2018; Alaoui et al., 2020). In repeated games where cooperation is possible, the level of intelligence of players is a crucial factor in affecting mutually beneficial behavior. Proto et al. (2019) find that when subjects are allocated into two groups on the basis of their intelligence, only the higher intelligence groups converge to full cooperation in complex, non-zero-sum games such as the repeated Prisoner's Dilemma.

However, such separation of individuals into distinct classes of intelligence seldom occurs in everyday life. The experimental literature has shown that when people are not separated by level of intelligence, they tend to choose efficient strategies leading to cooperation under repeated interactions, when gains from cooperation are sufficiently large (e.g. Dal Bó, 2005; Dreber et al., 2008; Duffy and Ochs, 2009; Dal Bó and Fréchette, 2011; Blonski et al., 2011). The questions of how groups interact and how people of different types within that group influence one another remain open.

To clarify these issues, we conduct an experiment allowing to directly compare how people cooperate when they are repeatedly interacting with individuals of the same or different levels of cognitive skills. In our laboratory experiment we find strong evidence that low IQ players learn to cooperate from high IQ players. Cooperation rates in integrated groups – where subjects with different IQ levels are pooled together– substantially increase among less-intelligent players (with IQ in the 76-106 range), and slightly decrease among more-intelligent players (those with IQ in the 102-127 range) when compared to rates in groups separated according to IQ.

We also find evidence of systematic differences in the strategies used in the integrated-intelligence setting, as compared to those in the separated-intelligence setting. High-IQ players in the integrated treatment are less likely to revert to cooperation after a defection of the partner than they do when playing in the separated treatment. To better understand the complex interaction of learning and teaching that produces this shift, we analyze a model that introduces differences among players in the ability to consistently and precisely implement strategies.

Before illustrating this model we need to recall some definitions, relying on an established classification of repeated game strategies, in particular for the Prisoner's dilemma (see for example: Aoyagi et al., 2019). A strategy is called *lenient* if it does not revert to defection after a single deviation; *forgiving* if it returns to cooperation after a punishment of a defection. A strategy is called *cooperative* if it plays cooperation at some stage. Always Defect is not cooperative, Grim Trigger is cooperative, but not forgiving, and Tit-for-Tat is cooperative and forgiving but not lenient, Always Cooperate is lenient, forgiving and cooperative. The combination of those three criteria produces an order, which if restricted to the four strategies we mention is complete. We call this an order of *harshness*: Always Defect, Grim Trigger, Tit-for-Tat and Always Cooperate are in order of decreasing harshness. We find that if the fraction of players with limited cognitive skills in the population increases, and, thus, the average probability of errors also increases, players will adopt harsher strategies.

We model a strategy as an automaton; management of a strategy consists in correctly choosing the next state in the automaton, given the current state and the observed action profile, and then correctly choosing an action. Thus, possible errors are errors in transition and errors in action choice. The distinction between the two types of errors is important for data analysis but most importantly as a conceptual distinction.[1] The literature on refinement of Nash equilibria (started with Selten, 1975) has introduced errors as an essential component of the analysis, and those are naturally interpreted as errors in action choice. The distinction between the two types of error makes an interpretation of observed behavior easier. An error in transition is likely due to a fault in cognitive skills and, in particular, working memory storage of the strategy adopted, its current state, and updating the state given the observed actions; an error in action is likely due to lack of care and attention in choice.

We study the effect on the frequency of strategies in the population at the evolutionary equilibria of different benchmark models (the proportional imitation model and the best response model; Gilboa and Matsui, 1991; Schlag, 1998) and in a rational learning model. This is a natural modeling choice because in our experiment each individual adapts strategies to the environment by observing the behavior of randomly met opponents. The cognitive skills distribution determines the probability of errors, which in turn affects the implementation of the strategy. Higher probability of errors lead subjects to shift towards harsher strategies, in particular the Always Defect strategy.

In our data analysis we follow the hypotheses of the model and estimate the probability of errors and the equilibrium strategy. We find a pattern consistent with the predictions of the model: in subject pools where subjects commit more errors, players employ harsher strategies. This is the opposite of what happens when errors in detection are introduced (as in models of imperfect monitoring).[2] In this case experimental evidence suggests that the presence of errors in detection produces a shift in the direction of more lenient strategies (see for example Aoyagi et al., 2019). A systematic comparison between these two errors is an interesting topic of future research.

[1] We develop an algorithm which exploits this difference in section E of the online appendix.

[2] This observation was suggested by a clever referee.

Our results allow to establish a clear link between intelligence, errors and strategic choices in repeated Prisoner's Dilemma. This is important for several reasons. First, it provides further insights in the way cooperation arises and evolves among self-interested individuals. Second, it provides a framework linking limited cognitive capacities and strategic errors. Third, it can help to understand strategic interactions in environments typically characterized by high heterogeneity in cognitive abilities, thus allowing a better evaluation of alternative integration policies.

Our study is part of a broader research program trying to identify the role of personality traits in producing strategic behavior. A large literature in behavioral economics has emphasized in the last decades the role of individual differences in attitudes toward others. These theories suggest that social preferences, such as trust and altruism (facets of Agreeableness, one of the Big Five personality traits), as determining behavior in strategic environments (see e.g. Fehr and Schmidt (2006) for a survey). Jagelka (2019) uses factor analysis to study the connection between preferences and cognitive skills on the one hand, and choices on the other. His model highlights the link between cognitive skills and inconsistencies of choices in a static setup, not in a strategic repeated environment.

Here we explore a completely different mechanism based on cognitive abilities. To the best of our knowledge, our model is the first to analyze the relationship between probability of errors and strategic choices, where individual cognitive skills determine frequency of errors. Heterogeneity in strategy choice in this class of games has been well documented, yet we know very little about why subjects do not coordinate on the same strategy or what determines the strategies selected, especially in a basic environment of perfect monitoring. Aoyagi and Fréchette (2009) estimate how the threshold strategies change as a function of the quality in monitoring. Fudenberg et al. (2012) show that if the action choice of players is implemented with an error by the experimenter, then strategies become more lenient and forgiving. So when subjects know that the error is not the responsibility of partners they might change their attitude in a more lenient direction, possibly because punishing the partner could not change the probability of error by inducing them to be more careful. Dal Bó and Fréchette (2019) report that as supergames become longer in expectation, the ratio of Tit-for-Tat to Tit-for-Tat plus Grim Trigger increases. There is also evidence that players in laboratory settings teach others how to play efficiently (e.g. Camerer et al., 2002; Hyndman et al., 2012; Cason et al., 2013). However, these papers do not connect the experimental findings with differences in intelligence and do not provide a unifying theory as we do in the current work. Furthermore, the explanation given in this paper based on evolutionary models relates to multiple observations from prior research: i) the importance of dynamics within sessions; ii) the predictive role of the initial conditions; and iii) the decrease in strategy implementation errors with experience (see Dal Bó and Fréchette, 2018, for a survey of such findings).

The paper is organized as follows: Section 2 presents the experimental design. Section 3 shows the experimental evidence. Section 4 describes the main results of a model of evolutionary game theory; the model is examined in more detail in the appendix (section A). Section 5 estimates the main parameters and variables of the model using our experimental data and provides support for the model's main assumptions and predictions. Section 6 concludes. The online appendix contains technical analysis of the model, estimation details, robustness checks, further experimental design details, and descriptive statistics.

2 Experiment Design and Implementation

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, which is measured during the first part. They are then 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: the first treatment (called IQ-Separated treatment, or *separated* for short) separates subjects into two groups, one with high and one with lower cognitive ability. The second treatment (called IQ-Integrated treatment, or *integrated* for short) pools together subjects into groups that do not differ one from the other in the distribution of cognitive ability. Subjects were not informed how the grouping was made. During the de-briefing stage (after the completion of the second day session), we ask the subjects if they understood the basis upon which the grouping was made. Only one subject mentioned intelligence as the possible determining characteristic.

2.1 Experimental design

Day One

On the first day of the experiment, subjects complete a Raven Advanced Progressive Matrices (APM) test consisting of a 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 by selecting one out of eight possible choices presented on the screen. Before beginning the test, the subjects are shown an example of a matrix together with the correct answer for 30 seconds. The 36 questions are presented in order of progressive difficulty (as sequenced in Set II of the APM). Participants are allowed to switch back and forth through the 36 questions during the 30 minutes and they 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-than-average intelligence. The test is able to elicit stable and sizeable differences in performances among individuals similar to the ones in our subject pool. This test was among others implemented in Gill and Prowse (2016) and Proto et al. (2019) that find it to be relevant in determining behavior in cooperative or coordinating games.

Psychometric and experimental economics research seldom offers subjects rewards for completing IQ tests such as the Raven. However, some research (e.g. Larson et al., 1994) has found that offering a monetary reward leads to a slight increase of scores among people of higher-than-average intelligence. Alaoui and Penta (2018) study the effects of such incentives on performance from a theoretical point of view. We aimed to measure intelligence in a way that would keep potential confounding with motivation at a minimum. Thus, we decided to reward our subjects with 1 euro for the correct answer to three randomly chosen matrices among the total 36. During the session we never mention that this is a test of intelligence or cognitive abilities.

Following the administration of the Raven test, subjects complete an incentivized Holt-Laury task (Holt and Laury, 2002) to measure risk attitudes. Participants also complete a standard personality traits questionnaire, the Big Five Inventory (BFI), which includes 44 questions (with answers coded on a Likert scale).[3] Participants also answer a series of questions on demographic characteristics, subjective well-being, and previous experience with a Raven's test. We inform the subjects that they will not receive any monetary payment for this section. The online appendix includes all instructions given on the first day.

Day Two

Subjects return to the lab on a later day for the second part. We allocate subjects into two experimental sessions depending on the treatment. In the *IQ-Separated* treatment, we divide subjects into two groups. Those with above-median Raven scores participate in what we refer to as *high-IQ* sessions; the remaining subjects participate in what we refer to as *low-IQ* sessions.[4,5] In the *IQ-Integrated* treatment, we create groups of similar Raven scores across pairs of sessions. To allocate subjects in second-day sessions, we rank them by their Raven scores and separate the groups by the median. We then alternate allocating subjects

[3] We use the version that was developed by John et al. (1991) and was investigated by John et al. (2008).

[4] The attrition rate was small. Table O.1 in the online appendix provides the details.

[5] In cases in which there were subjects with equal scores at the cutoff, a tie rule based on whether they reported previous experience of the Raven task was used. Participants who had done the task before (and were tied with others who had not) were allocated to the low-IQ session.

of either intelligence level in different sessions.[6] Specifically, the two top Raven scorers of day one are allocated in different sessions; subsequently the third and fourth are allocated to a different session, and so on.

Subjects are asked to play several instances of a repeated Prisoner's Dilemma (PD) game. Table 3 reports the stage game implemented. We use a random continuation rule: after each repetition the draw of a random number determines whether the repeated game stops or continues to the next repetition. The continuation probability is $\delta = 0.75$. We use a pre-drawn realization of the random numbers; this ensures that all sessions across both treatments face the same experience in terms of length of play at each decision point.[7] We define key terms as follows: *round* refers to an overall count of number of times the stage game has been played during the session; a *supergame* refers to each repeated game played; and, finally, *period* refers to the round within a specific supergame. The length of play of the repeated game is either 45 minutes or until the completion of the 42nd supergame (which entailed 151 rounds, i.e. repetitions of the stage game), depending on which comes first. This rule was not disclosed to the participants. The game parameters are identical to the ones used by Dal Bó and Fréchette (2011) and Proto et al. (2019). When the continuation probability is interpreted as discount factor, the payoffs correspond to an infinitely repeated Prisoner's Dilemma game in which the cooperation equilibrium is both subgame perfect and risk dominant.[8]

The matching of partners takes place within each session under an anonymous and random re-matching protocol. Two participants play as partners as long as the random continuation rule determines that the particular partnership is to continue. Once a match ends, participants are again randomly and anonymously matched, and start playing the game again. Each decision round for the game is complete when every subject has made their decision. After all subjects make their decisions, a screen appears that reminds them of their own decision, indicates their partner's decision, and reports the units they earned for that particular round. The group size of different sessions varies depending on the numbers recruited in each week.[9] The subjects are paid according to the points they earn through all rounds of the game. The payoffs reported in Table 3 are in terms of experimental units; each experimental unit corresponds to 0.004 Euros.[10]

[6] Again, the attrition rate was small. Table O.2 in the online appendix provides the details.

[7] In table O.8 in the online appendix, we list the length per supergame that was used for all sessions. This realisation of supergame length was obtained by drawing a random sequence.

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

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

[10] In the invitation email sent to potential participants we announced a maximum on earnings of 26 Euros for the entire study (including a 4 Euros participation fee). The invitation letter for the second day

After the sequence of repeated games is completed, subjects are asked to complete a short questionnaire about any knowledge they have of the PD game, to answer some questions related to their attitudes towards cooperative behaviour and some strategy-eliciting questions. We include all these questions in the online appendix.

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 took place 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, including 4 Euros for participating. The experiment used *Z-Tree* (Fischbacher, 2007).

We conducted eight sessions for the separated treatment: four high-IQ and four low-IQ sessions. There were 108 participants, with 54 in the high-IQ and 54 in the low-IQ sessions. For the integrated treatment we conducted 8 sessions with a total of 106 participants. Tables O.1 and O.2 in the online appendix report the dates of the sessions and the number of subjects per session.

3 Experimental Evidence

3.1 Cooperation rates, payoffs and learning

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 first period choice across all treatments. Subjects in the high-IQ sessions converge faster to almost total cooperation rates, while in the low-IQ sessions the pattern emerges more slowly; subjects in this group converge to a cooperation rate smaller than 100 per cent (left panel). Subjects in the integrated session have higher cooperation rates than in the low-IQ sessions, and lower cooperation rates than in the high-IQ sessions (right panel).[11] Always from figure 1, we note a clear upward trend only in the 1^{st} half of each session (i.e. about the first 20

announced a maximum 15 Euros to be earned during that second session (up to 7 plus 4 Euros could be earned during the first day). During the second day session, no display of the accumulated amount earned up to that point was shown on the screen to the participants. The final earnings for the whole experiment were only communicated at the end of the experiment. The results that we report are not qualitatively affected by excluding the data in every session when even a single subject has reached the maximum on earnings, so the presence of the maximum earning did not affect the behavior of participants.

[11] In Figure O.2 of the online appendix, we present the cooperation rates by session.

supergames), for this reason on what follows we will generally show separate estimations for the two halves of every session.

Table 4 shows that in the first 20 supergames (roughly the first half of a session), subjects earn about 2.5 units more, and cooperate about 10% more frequently in the high-IQ sessions than in the integrated sessions. By contrast, they earn about 5.6 units less and cooperate about 22% less frequently in low-IQ sessions than in the integrated sessions. After the 20^{th} supergame, there is no longer a significant difference between high-IQ and integrated sessions. Meanwhile in the low-IQ sessions the difference in both payoffs and cooperation remains constant.[12]

In accordance with the findings in Proto et al. (2019), Table 5 shows that IQ is not significant in determining cooperation in the first round in either of the two treatments. This suggests that the difference in cooperation between individuals of different cognitive skills is only due to a learning effect during the session.[13] Furthermore, in columns 3, 4 and 5 of table 5 we include a dummy for the integrated treatment. As it is expected it is non significant, confirming that subject behaviour is statistically indistinguishable across treatments at the beginning of the session.

Figure 2 shows that the average payoff per supergame is consistently higher in the integrated sessions than in the separated sessions, indicating that in the integrated treatment subjects play on average more efficiently than in the separated treatments.[14] Figure 2 suggests that subjects learn to cooperate in the first half of each session, then cooperation rates become stable in the second half.

Table 6 shows that – in the first half of the session – subjects in the high-IQ sessions open with cooperation at a rate increasingly faster than in the integrated treatment. By contrast, cooperation increases more slowly in the low-IQ sessions than in the integrated sessions (columns 1 and 2). Subjects in the low-IQ sessions tend to catch-up with the others in the second part of each session (columns 3 and 4). This provides evidence that the less intelligent players learn to play more cooperative strategies more quickly when they mix with the more intelligent players than when they play with players of similar intelligence level.[15]

[12] In table O.16 of the online appendix we present the same estimation as in table 4 but using only the data from the marginal subjects in terms of IQ by excluding the top and bottom tertiles. This gives perhaps a cleaner estimation of the magnitude of overall treatment effect. The table shows that there is an even bigger difference between low IQ sessions and combined sessions (30% less cooperation in the latter), while the difference between high IQ sessions and combined seems smaller than in table 4.

[13] Interestingly, risk aversion is the only significant negative determinant of cooperation at the beginning of each session; in Proto et al. (2019) the effect was negative as well, but not statistically significant.

[14] Table O.3 in the online appendix also underscores this trend, with total earnings and average payoff per round significantly higher in the integrated sessions than in the separated sessions.

[15] In table O.17 of the online appendix we present the same estimation as in table 6 but using only the data from the marginal subjects in terms of IQ by excluding the top and bottom tertiles. The table shows

Why do higher levels of cooperation emerge earlier in the session with a larger number of high-IQ players? One driver appears to be subjects' beliefs, which change on the basis of the experience with previous partners' opening plays. As columns 2 and 4 of Table 6 show, partners' past cooperation in the first periods significantly increases the probability of cooperation; in agreement with established results in the repeated games literature. In other words, subjects whose partners opened with cooperation in previous supergames are more likely to open with cooperation in subsequent games.

It is reassuring to observe that the results so far confirm the results in Proto et al. (2019) with the same payoff matrix, but using a different subject pool in a different country.[16] We summarize these findings as follows: *Average cooperation rates and aggregate payoffs are higher when high- and low-IQ subjects play in integrated, rather than separated groups. Low IQ individuals learn to cooperate faster when they play in integrated, rather than separated groups.*

In what follows we will ask and study the fundamental question of how the interaction between IQ distributions, strategy choice and implementation errors materializes.

3.2 Individual choice determinants

There is widespread and robust evidence (e.g. Dal Bó and Fréchette, 2018; Proto et al., 2019) that subjects in the repeated Prisoner's Dilemma game with perfect monitoring play *memory one* strategies, where the choices in every period of a supergame depend only on the action profile of the previous period.[17] Let $ch_{i,t}$ represent the subject's 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 time-invariant individual fixed-effects (taking into account time-invariant characteristics of both individuals and sessions) and $\epsilon_{i,t}$ is the error term. We estimate Model 1 separately

that there is an even bigger difference between the cooperation trend of low IQ and combined sessions.

[16] And are generally consistent with the findings in the experimental economics literature (see e.g. Dal Bó and Fréchette, 2018)

[17] Common examples of memory one strategies are Tit for Tat or Grim trigger. Defecting after two periods of defection of the partner (as in Tit for Two Tats) illustrates a memory two strategy.

for the first block of 20 and the second block of 20 supergames by using a logit estimator.

In Table 7, we present the estimates of Model 1 for the first and the second half of the session. Results are reported in odds ratios and taking the outcome $(C, C)_{t-1}$ as baseline. In panel A, we note that, conditional on deviation from mutual cooperation (i.e. after $(D, D)_{t-1}$; $(D, C)_{t-1}$; $(C, D)_{t-1}$), the odds of cooperating at time t by high-IQ are higher when subjects play with other high-IQ players than when they play in the integrated treatment. This difference is, if anything, even larger in the second part of the session (as seen in panel B of Table 7). In particular, a lower coefficient of $(C, D)_{t-1}$ suggests that subjects are less lenient (see the definition in the introduction), and a lower coefficient of $(D, C)_{t-1}$ suggests that subjects play less forgiving strategies.[18] Overall, we can interpret this evidence by saying that in the integrated sessions high-IQ adopt *harsher* strategies.

Table 8 shows results of a direct test on whether high-IQ players are less harsh when they interact amongst each other than when they are in the integrated treatment.[19] We note that the high-IQ players are significantly less likely to cooperate whenever the other subject unilaterally defects. By contrast, the low-IQ subjects do not seem to play significantly differently – regardless of the setting, whether playing only with other low-IQ or in the integrated treatment after an unilateral defection. We summarize this section as follows: *High-IQ players are significantly less likely to cooperate after an unilateral deviation of the partner when playing in the integrated treatment than when playing in the separated treatment.*

The evidence we report shows that subjects learn to cooperate, and their choice of strategies depends on their own cognitive skills as well as the distribution of cognitive skills within the group. The high-IQ players seem to be more lenient when they play only with other high-IQ players than when they also play with low-IQ players. Therefore, a complex mechanism is in place and the following model allows us to analyze this in detail.

4 Errors and Evolution of Strategies

Proto et al. (2019) found that cognitive skills are inversely correlated to the probability of defaulting after mutual cooperation in a strategic dynamics setting, suggesting the existence

[18]However, such a coefficient could also be an indication of reverting back after an error at period $t-1$. In a similar vein, the interpretation of a lower coefficient of $(D, D)_{t-1}$ is more ambiguous: it might suggest that a player is less keen to re-establish cooperation, that is, they are playing an non-cooperative strategy like Always Defect or a Grim-trigger. At the same time this might also suggest, to the contrary, that players adopt Always Cooperate more often and are reverting to that after an error at period $t-1$.

[19] Given our between-subjects treatments, we could not estimate this model by using individual fixed effects to control for individual heterogeneity. As a result, we needed to introduce correlated random effects in the model; a probit estimator is a more correct choice in this case (Wooldridge, 2010, sections 15.8.2 and 15.8.3)). However, note that the logit estimator would still be a good approximation, and using it instead would give qualitatively similar estimations.

of a link between cognitive skills and strategic mistakes. Our results rely on a precise model and show support for this suggestion. In this section we precisely define two types of errors that take place in our environment and describe a formal model of the role that the probability of these errors have on the evolution of strategies. We assume that error rates are related to the intelligence of players, particularly errors that are due to ineffective information processing: thus, the model provides a link between intelligence and strategic behavior in groups. In section 5, we provide empirical evidence of this assumption in our data.

4.1 Two types of error

In our model of strategy evolution, subjects can make two types of errors when implementing their strategies. As a first possibility, errors may occur 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 as a consequence of the observed action. We model subjects' plans as automata, and the complex process we have just described as the transition to the next period state in their automaton. An error produces a transition to a state which is different from the one prescribed by the automaton and the observed sequence of actions. We call this an *error in transition* and denote its probability of occurring in each instance by ϵ_T. We also consider the second possibility that they make an error in their choice of action, *error in action* for short, which is the selection of an action different from the one prescribed by the strategy at the current state. This error occurs with probability ϵ_A.

Note that the error in action is the only type of error which is possible when a player uses simple strategies such as Always Defect or Always Cooperate, in which the automaton has a single state. Thus, when we estimate error rates in our data analysis, if the strategy considered is, say, AD, and only errors in transition are allowed, an observed path of actions of a player has either probability one (when only defection is observed) or probability zero (if ever a cooperation is observed). Thus, allowing this second type of error is necessary to allow the possibility of errors with simple strategies such as Always Defect. We believe that the distinction between these two types of error is important, because it corresponds to different processes and different individual characteristics. From the cognitive point of view, an error in transition is typically an error in working memory implementation of the strategy; an error in action choice is an oversight, and more likely to be induced by distraction or lack of focus. This distinction is likely to be even more important in future research investigating the link between individual characteristics of players and strategic behavior.

4.2 Errors and Strategy evolution

We now examine in broad lines the link between the frequency of errors and the evolution over time of strategic behavior. For clarity of exposition we focus in the main text on the simple case of transition errors only, and extend the analysis in section A.3 of the appendix, where we consider the general case in which both errors in action choice and in transition are possible. Focusing on transition error considerably simplifies the exposition, and the results in section A.3 show that this case contains all the essential features. Our data analysis tests the complete model with both types of errors.[20] In section A.4 of the appendix we present a model of population learning that allows us to establish a link with social learning models of the type used in Dal Bó and Fréchette (2011).

We use the model to study the effect of error probabilities on the frequency of strategies adopted by players. It is useful to have in mind the order of magnitude of error probabilities that is actually observed in our data.[21] The range of error probabilities in our data is for almost all participants within the range of 0 to 0.3. The mean is between 0.02 and 0.1, but Figure O.5 (in the online appendix) shows that in earlier supergames the mean can be much higher, up to 0.25 or even larger. Overall, it seems reasonable to focus our attention conservatively (to consider all possible relevant cases) to the range 0 to 0.25.

4.3 Sequences of Repeated Games

In our experiment and in our model, subjects play consecutively several instances of a repeated game, with a different partner randomly chosen in each repeated game. The two-players stage game with action set A^i for each player, $A \equiv A^1 \times A^2$, of each repeated game is a Prisoner's Dilemma with payoff in every period given by:

Table 1: Stage Payoff u

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

where we assume:

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

[20]The estimation method is presented in section E of the online appendix.

[21]In Figures O.3 and O.4 in the online appendix, we report the histograms of errors in transition and action, respectively, by treatment. The dispersion of transition error probabilities across individuals is reported in Figures O.5 (for early supergames) and O.6 (for all supergames).

These conditions are satisfied by the payoffs in our experimental design ($c = 48$, $s = 12$, $t = 50$, $d = 25$). We denote by u the payoff in Table 1, dropping the index using the symmetry of the payoff matrix.

In the following an automaton is a representation of repeated game strategies of a player. An automaton m is a tuple (X, x_0, f, P) where X is the set of states of the automaton, $x_0 \in X$ 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 row or column player who is using that automaton, relying on the symmetry of the game. Finally, P defines the transition probability $P : X \times A \to \Delta(X)$, and write:

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

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 case, considered later, in which we introduce errors.

We assume that in the repeated game subjects only play one of the following three strategies: Always Defect, Tit-for-Tat and Grim Trigger strategies. This assumption relies on by now established findings in the experimental study of the repeated Prisoner's Dilemma, which are replicated in our own data.[22] In the data analysis of section 5 we also consider the possibility that some players use the strategy Always Cooperate (AC). Since this strategy is weakly dominated and with errors would be eventually eliminated from the population, a model including AC would give conclusions similar to those we reach here. To lighten notation, in the current section we denote these strategies as $\{A, G, T\}$, where A stands for *Always Defect* strategy, G for *Grim Trigger* and T for *Tit-for-Tat*. Note that in our description of strategies as automata, there is a single state for A, and two states for G and T.

We can use conceptual tools from evolutionary and learning game theory in our setup, in which a game is a repeated game, by considering the normal form game where the strategy set for each player is the set of strategies or equivalently automata:

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

The normal form game is defined as follows: before the beginning of each repeated game, players choose simultaneously an element in M, and the payoff is the one induced in the

[22] For example, Dal Bó and Fréchette (2019) show that subjects mostly adopt these strategies. Dal Bó and Fréchette (2011) and Proto et al. (2019) find evidence that the frequency of use of these strategies by their participants ranges between 66% and 90%.

repeated game by the chosen pair of strategies. At the end of each repeated game they play another one, by again choosing a possibly different repeated game strategy, and so on: so players play a sequence of repeated games. We call the normal form game induced by the choice of strategies in a repeated game the *strategy choice* game. We describe now the payoffs when the set of strategies is as in (4).

When strategies are implemented by both players with no error (so $\epsilon_T = \epsilon_A = 0$), the precise values of the payoff matrix are reported for convenience in table 2.

Table 2: **Payoff Matrix in the Strategy Choice Game with No Errors**.

	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 denote by U the payoff in Table 2. When one of the two error probabilities, or both, are strictly positive then the matrix assigning payoffs to each pair of strategy choice by the players changes. For example, if $\epsilon_T > 0$ and $\epsilon_A = 0$, then the payoff from the choice of grim trigger by both players is no longer (c, c): with probability ϵ_T in every round a player will transit to the wrong state, choose defect, thus inducing retaliation of the other. The computation of the payoff matrix corresponding to the simple Table 2 when errors are possible is slightly more complex, but it produces a payoff matrix similar to Table 2, where entries are a function of the pair (ϵ_T, ϵ_A). In the appendix section A.2 we describe precisely how this new payoff matrix is computed when $\epsilon_T > 0, \epsilon_A = 0$. The general case is analyzed in section A.3.

4.4 Evolutionary Models

We study the effect on the frequency of strategies in the population for different values of the error rates at the evolutionary equilibria of two models: the *proportional imitation model* (Schlag, 1998), which produces the replicators dynamics equation described in equation (5) below, and the *best response model* (Gilboa and Matsui, 1991), which produces the best response dynamics described in equation (6) below. Finally we consider a rational learning model (see section A.4 for details), which is very close to the best response model, and is also a realistic description of learning in our setup of consecutive repeated instances of a repeated game with randomly matched opponents.

There are several reasons to consider each of these models: each one has its merits. First, the replicator dynamics is the most widely used in evolutionary game theory, and it is therefore a natural benchmark. Its behavioral foundation, proportional imitation, does not require the knowledge of the strategies' distribution by the player, who observes after the fact the partners' strategy and payoff, and chooses whether to imitate it or not.[23] But the behavior assumed (imitation of a more profitable strategy) is crude; the assumption that the player gets to know the strategy the other played is not natural in our setup, where the game is the normal form of a repeated game. The best response dynamics is based on a more realistic behaviour in our environment, but it requires players to know at any time the distribution of the strategies in the same experimental session. A model which is useful and a reasonable description of behavior in our experimental environment is the model of rational learning that we present in section A.4 of the appendix. This model assumes that players have a belief over the strategy distribution of the subjects in the session, update it on the basis of the observation of the partner's behavior, and choose an optimal response at each time. This model allows us to predict behavior during the learning process, and not only at the steady state. The price of this model is its complexity: so the other two, proportional imitation and best response, are very useful to provide an introductory tractable setup. They also provide a useful approximation because (as we show) the qualitative results are robust to different specifications of the evolutionary model (see Proposition A.2), in particular the main conclusion, that harsher strategies are more likely when the average error probability is higher is common in all three models.

We now present the details. We let $\mu \in \Delta(M)$ denote a mixed strategy and also a frequency of choice of strategy in the population (as usual: see Sandholm, 2010; Weibull, 1997). To describe the evolution over time of the frequency, we let $\mu(t, m)$ denote the frequency of strategy m in the population at time 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 $\tau \in \Delta(M)$, the average payoff in the population as:

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

The time evolution of the frequency under proportional imitation is the replicator dynamics equation:

$$\forall m \in M, \frac{d\mu(t, m)}{dt} = \mu(t, m) \left(U(m, \mu(t, \cdot)) - U(\mu(t, \cdot), \mu(t, \cdot)) \right). \tag{5}$$

A model of matching and imitation justifying the proportional imitation in (5) is standard (Sandholm, 2010). Equation (5) states that the change in a short time interval of a frequency of a strategy is proportional to the difference between average payoff from that strategy and

[23] See Schlag (1998) and Sandholm (2010) for more details.

the overall average payoff in the environment, times the current frequency of the strategy.

To define the best response model, we first need to define the best response correspondence, taking values in mixed strategies:

$$BR(\mu) \equiv \{\tau \in \Delta(M) : \forall m \in M, U(\tau, \mu) \geq U(m, \mu)\}. \tag{6}$$

Let $\lambda > 0$ be a fixed parameter, controlling the speed of adjustment. The time evolution of the frequency under best response dynamics [24] is described by any solution of the differential inclusion:

$$\forall m \in M : \frac{d\mu(t,m)}{dt} = \lambda \left(\tau(m) - \mu(t,m)\right), \text{ for some } \tau \in BR(\mu(t,\cdot)). \tag{7}$$

To clarify this equation we note that $\mu(t,\cdot)$ is a probability measure on M; the best response correspondence BR is then applied to $\mu(t,\cdot)$. If for a time t and an element $\tau \in BR(\mu(t,\cdot))$ equation (7) holds, then the function of time described by $\mu(t,m)$ is a solution at that time. In summary, a solution of the best response model is an absolutely continuous function for which equation (7) holds at every time t.[25] This model is simpler than it may look. Figure 3 describes the dynamics in the best response model, and illustrates that the analysis is simplified in our case because when the error probability is strictly positive the best response correspondence is single valued in the interior of each the three basins of attraction. Thus, the differential inclusion is, for "most values", a simple differential equation. Equation (7) states that at every point in time the frequency of the strategy which is the best response to the current frequency increases, that of the other decreases, at an overall speed regulated by the parameter λ. Since the best response dynamics is closer than the proportional imitation dynamics to our environment, we provide for it a justification in the next section.

4.5 A Model of Adjustment of the Best Response

A population of players (like the subjects in our sessions) know the distribution μ and can compute the best response, but do not adapt immediately the best response. For example, they may need time to think through the best response, or can pay attention to the current situation only at some prescribed time. In a small time interval dt a fraction λdt of the population revises their strategy; the complement $1 - \lambda dt$ does not. Those who do, adapt the best response to the current $\mu(t,\cdot)$; each player does so taking the current μ as given,

[24] We follow Sandholm (2010)'s formulation.

[25] See Aubin and Cellina (1984) for technical details on differential inclusions; in particular chapter 2, section 1 on differential inclusions where the correspondence is convex valued and upper-hemicontinuous, which by the properties of the best reply correspondences for normal form games is the relevant case for our analysis.

and ignoring (correctly, since he is negligible) his effect on the frequency of strategies. So for every m we have:

$$\mu(t + dt, m) = (1 - \lambda dt)\mu(t, m) + \lambda\, \tau(m)\, dt, \text{ for } \tau \in BR(\mu(t, \cdot)), \quad (8)$$

which in the limit gives the differential inclusion (7). In this version of the model, players in a large population know exactly, at every point in time, the frequency of strategies in the population; they just cannot or do not adjust immediately. 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. It differs from a full rational learning model because it requires that all players have the same belief and the belief is correct. The exposition of results for the rational learning model is harder because a simple two-dimensional dynamic is impossible. The rational learning model is defined in appendix section A.4, and estimated in section A.5.

4.6 Summary of Main Results

We now outline our main results. As we mentioned already, the reader may focus on the case in which $\epsilon_A = 0$, that is there are no errors in action, and the only error possible is error in transition.

First, the limit behavior of the fraction of strategies is determinate (that is, the steady states are locally unique) only if the probability of error is positive.[26] When there are no errors, typically the equilibria of the strategy choice game, as well as the steady states of the evolutionary model and of the learning process, are not locally unique. The unique exception is 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 strategy equilibria of the game in which players choose strategies in the repeated game. With small errors, 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. The basin of attraction of a strategy is a useful quantitative indicator of how profitable on average a strategy is, given a specific environment (which includes payoffs and error probabilities). It may also provide a theoretical basis to predict relative frequency of strategies as a function of the error probability. The link between frequency of the strategy and basin of attraction is provided by the following thought experiment. The initial condition of the system is the initial distribution

[26]The role of mistakes in evolutionary stability has been explored, extending the research on Nash refinements, by Boyd (1989).

of strategies in the population. We may think that this is chosen randomly with a fixed probability, that is, independent of the environment, in particular of the error probability and continuation probability. For example, to fix ideas, we can take this probability as the uniform probability over the strategies' frequencies. This probability may represent the random factors affecting the initial behavior of subjects in an experimental session. Given an initial condition, the evolution of distribution of strategies and its limit frequency is then provided by the evolutionary model. If we keep this initial probability fixed, and we change the model's parameters, in particular the error probability, a larger basin of attraction of a strategy will be associated with a larger frequency of that strategy. Thus, the size of basin of attraction gives us a measure of the probability that the final distribution of strategies is A or G. The size of these basins changes with the probability of errors. As players become more error prone, the basin of attraction of harsher strategies become larger: that is, the size of the basin of the A strategy becomes larger than that of G and T together, and that of G becomes larger than that of T.

Third, which strategies survive in the long run depends on the discount factor δ; in all cases, for low probabilities of errors all strategies may survive depending on the initial condition. For high probabilities of error, only A survives. For intermediate probabilities, the two surviving strategies are A and G for low δ, and A and T for high δ. As δ tends to 1, keeping error probabilities fixed, the opposite happens: the basin of attraction of the G and T strategies becomes larger; and when strategies are limited to $\{G, T\}$, the basin of T increases to cover the entire interval. This is in line also with the experimental finding in Dal Bó and Fréchette (2019) that as the probability of continuation increases subjects tend to choose shorter punishments (i.e. T over G).

The results described so far are independent of the specific model of evolution we adopt, (Proportional Imitation or Best Response). In Section A.4 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. We show that the resulting dynamics are close to those described by the best response dynamics.

4.7 An Illustration

Figure 3 provides a first intuitive understanding of the way in which the error probability affects the basins of attraction of the three strategies.[27] The figure reports the phase portrait of the vector field for the Best Response dynamics described in equation (7). We set in all panels the probability of error in action $\epsilon_A = 0$, and we considered different error probabili-

[27]The Matlab code producing the analysis reported in this figure and those considered later is available upon request.

ties, ranging from $\epsilon_T = 0$ (no error, top left panel) to $\epsilon_T = 0.25$ (bottom right panel). Payoffs in the stage game are set as in our experimental design in Table 3. The discount factor (or equivalently, the continuation probability) is the same (0.75) as in our experimental design.

The triangle in each panel of figure 3 is the two-dimensional projection of the simplex. Each point in the triangle represents points $(\mu(G), \mu(T))$, where $\mu(G)$ is the probability of G, such that $\mu(G) + \mu(T) \leq 1$, and so the frequency of the strategies, in the order (A, G, T), is equal to $(1 - \mu(G) - \mu(T), \mu(G), \mu(T))$. The lines in the triangular regions represent the isoclines: an isocline is the set of points at which the time derivatives of a variable contains zero.[28] The red line indicates the set of points at which $\frac{d\mu(T)}{dt} = 0$, the blue ones the set $\frac{d\mu(G)}{dt} = 0$. These lines separate the triangular region into three subsets, each containing the point corresponding to a pure strategy.

The top right panel shows the case with a very small error probability ($\epsilon_T = 0.0001$). The two top panels illustrate the sudden change of the dynamic as soon as the error becomes positive although still very small. 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 triangular subset at the bottom left of the triangle. The fact that the boundaries of the regions are straight-lines follows from the special nature of the Best Response dynamics. This figure can be compared with Figure O.9 in the online appendix, reporting results for the Proportional Imitation model.

Several points that appear clearly in Figure 3 are worth pointing out. First, for all error probabilities, 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 the error probability we are considering in this example. Second, the size of the region attracted to A increases monotonically with the error probability. 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 converges to G (the strategy which is second in order of strictness) is increasing as the error probability is increasing. Note that as a result, whereas at the lowest error probability (top left) the T region is the largest, at the highest (bottom right) it is the smallest.

We can summarize this section by saying that *as the error probability increases, harsher strategies become more frequent.* The next section tries to explain the reason for the result.

[28]More precisely, since we define the Best Response dynamics as a differential inclusion, the set of points at which zero is in the set of derivatives. In the discretization used to produce Figure 3, the distinction between differential inclusion and differential equation is irrelevant.

4.8 Why Harsher Strategies Thrive with Larger Errors

When no error is possible, the strategies G and T produce the same outcome when matched with A, and the same outcome when matched with each other. So with no error both equilibria and dynamic behavior are indeterminate, because G and T are interchangeable.

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 A.4). Hence, no matter what the error is, the profile (A, A) remains a Nash equilibrium and a locally stable steady state. Thus, if some strategy is eliminated, it is not A. The second crucial fact is that at $\epsilon_T = 1/2$ under some condition on payoffs, which is satisfied by the payoffs in our experimental design, the pair (A, A) is the unique Nash equilibrium (see Lemma A.6). These two facts together with the continuity of the value function in ϵ_T (see Lemma A.3) imply that the set of strategies that survive change from the entire set to the strict strategy A. What is left to study is what happens during the transition.

With a small probability of error the value for (say) player 1 of using G when matched with G in the repeated game (call it $V(G, G)$) and the similarly defined values $V(T, G)$, $V(G, T)$, $V(T, T)$ 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 relative size of the basin of attraction between G and T changes. As previously mentioned, in the model with no errors the basin of attraction of G and T are not clearly defined. For infinitesimally small errors, the value of the steady state that separates 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 appendix A). 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 of error tending to zero) 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 50-50 probability of error, one can see that the only surviving strategy is the harshest strategy, A.

4.9 The Role of the Discount Factor

The effect of the discount factor on the basin of attraction follows naturally from the effects of δ on payoffs. Let us consider first the case of δ close to 1. Consider players starting from mutual cooperation. A small transition error can lead a player 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 of the player. Instead, the effects of such an error with Tit-for-Tat are not as durable. Hence the relative fitness of T increases, and its basin of attraction is larger.

The increase in the error probability ϵ_T reduces the comparative advantage of the forgiving 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 harsher strategy). For a graphical illustration of the role of the discount factor, compare Figure 3 (where $\delta = 0.75$) with Figure O.8 in the online appendix (where $\delta = 0.6$); in both cases payoffs are the same as in our experimental design.

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. What dictates the relative fitness of strategies is their relative profitability. In all the models, the effect of the frequency of errors is mediated by its effect on the payoff structure. This link is examined in detail in appendix A.

In the next section we test our model in our experimental data.

5 Errors and strategic choices in our data

In the previous section we have shown that in evolutionary game theoretic models, as the frequency of errors increases, players play harsher strategies more often. If we assume that the error probability is negatively correlated with cognitive ability, we should observe harsher strategies more frequently in the low-IQ separated sessions compared to the integrated, and also in the integrated sessions compared to the high-IQ separated. In this section we show that our experimental data provide support for the assumption on the relationship between error probabilities and intelligence, as well as for the predictions of the model. As we mention in section 4.3, the model considers the three strategies Always Defect, Grim Trigger and Tit-for-Tat, but the conclusions are easily extended to the case in which players may also choose Always Cooperate. In the data analysis below we allow subjects to use Always Cooperate.

Our analysis rests on an algorithm identifying the set of error histories consistent with the observed sequence of actions. The algorithm is described precisely in section E of the online appendix.[29] We assume that subjects choose a strategy at the beginning of a finite sequence of one or more consecutive supergames (that we will call *block* of supergames) and implement this strategy, possibly with errors (in action or in transition).

[29]The Matlab code computing the sequence of errors in action choice and transition consistent with the observed sequence of action choices, and that computing the maximum likelihood, is available upon request.

For each block of supergames we identify the strategy or strategies most likely to have been utilized given the decision sequence and implied possible error histories. We define the variable *ties* as the number of strategies having maximum likelihood. The number of strategies at the maximum ranges from 1 (and in this case the data allow us to precisely identify a single strategy) up to 4 where four strategies have the same likelihood. Thus, observed data do not always allow to precisely identify a strategy.[30]

In determining the block size to be used in the estimation we are faced with a trade-off. By increasing the block size, we are able to use additional information than can help disentangle across the strategies considered. Table O.20 in the online appendix shows that indeed the number of times the most likely strategy is unique (i.e. the variable *ties* is 1) decreases with block size. However, increasing block size is also costly for the estimation. By including more supergames within a block, we are more likely to pool supergames across which participants may have switched strategies. This causes a reduction on the likelihood obtained for all strategies. This is clear from figure O.10 in the online appendix, where we show the average likelihood of the most likely strategy across the session for varying block sizes. The average likelihood decreases with block size.

In the discussion of our analysis below we focus on the estimation results with blocks of five supergames. The analysis yields qualitatively similar results if we use alternative block sizes, three for example. Note how in figure O.10 of the online appendix, after the first third of blocks, the average likelihood is around 80 percent in all three panels. This suggests that subjects are overwhelmingly playing the four strategies that we consider in our model, as has already been noted previously (e.g. Aoyagi et al., 2019).

The assumption that higher errors are associated with lower intelligence is empirically tested in table 9.[31] In columns 1 and 4, we can observe that both error probabilities are negatively correlated with IQ. The correlation is significant at a higher significance level for the errors in transitions than for the errors in actions as it is perhaps natural to expect given our errors' definitions. In columns 2 and 5, where we control for length of play, the significantly negative coefficients for block of supergame indicate that experience reduces error probabilities; we will return to this effect in the next subsection. Finally, in columns 3 and 6, we introduce a dummy for integrated sessions, showing that this relationship is not significantly different in the separated and integrated treatments.

In Table 10 we present a summary of the results of these estimations, where all most likely strategies and the frequency of errors in transition and errors in action for each treat-

[30]For example, take a sequence of mutual cooperation (C^1, C^2) over the entire supergame. This sequence is consistent with three strategies, i.e. AC, TfT, GT, with no errors. In this case the likelihood for each strategy would be equal to one.

[31] The error estimation algorithm is described in section E of the online appendix.

ment group are reported. The lower the proportion of high-IQ subjects in the treatment is, the higher the probabilities of errors are (ϵ_T and ϵ_A in the model). The high-IQ sessions have a lower error frequency than the low-IQ sessions and integrated sessions and the integrated sessions have a lower error frequency than the low-IQ sessions. The error probabilities are significantly different between low-IQ separated sessions and integrated sessions. The comparison of the error rates between low-IQ in the separated and integrated sessions show that low-IQ individuals learn to commit fewer errors in action (last column of table 10), potentially because in this treatment they are more focused since the high-IQ players use more harsh strategies. We will elaborate more on this point below. The convergence of error probabilities between intelligence groups when mixed together is also evident from the comparison of high-IQ and low-IQ groups within the integrated sessions where no statistically significant differences appear (bottom two rows of column 'Diff. 4 - 5' in Table 10). Interestingly, there is little difference in the transition and action errors of the high-IQ subjects.

We now turn to the frequency of strategies depending on their harshness. There are significant differences in frequency among the various types of sessions. Sessions with more high-IQ subjects tend to feature less harsh strategies, high-IQ choose AC, TfT and GT more often than low-IQ, and AD less often. Similar tendencies are evident when comparing integrated sessions with low-IQ sessions: subjects in integrated sessions choose cooperative strategies (AC, TfT, GT) more often than in low-IQ sessions. The opposite is true, when we compare high-IQ sessions with integrated sessions: in this case subjects in high-IQ sessions choose cooperative strategies more often. In summary, we observe that a higher share of high-IQ subjects is associated with lower error frequency and less harsh strategies.

Comparing subjects with similar IQ across the different treatments, we note that High-IQ subjects are more lenient in the separate sessions by playing AC more often (column 'Diff 1 - 4' of table 10) than in the integrated sessions. On the other hand, low-IQ are significantly more lenient (more AC) and more forgiving (more TfT and less AD) in the integrated sessions.

When we have strategies GT and TfT tied (in terms of estimated likelihood), we cannot determine which of the two the subject is following. Table O.21 in the online appendix, shows that for blocks in which strategy GT is the most likely strategy, strategy TfT is also in 85 percent of the cases (943/1,107), and for blocks in which strategy TfT is the most likely strategy, GT is also in 88 percent of the cases (943/1,070). This is not of course a sign of weakness either of our estimation method or of our design: with no errors, and all players using GT or TfT, the two strategies would be the most likely strategies in all cases.

5.1 Dynamics of Errors over the Session

The model described in section 4 and appendix A, and its extension in section A.3 to two types of errors, assumes that the probabilities ϵ_T and ϵ_A, are fixed. In our data instead, these probabilities decrease as subjects gain experience, as suggested by the regression results reported in columns 2 and 4 of Table 9. In Table 11 we analyze and compare the first and the second part of the session for each treatment (consisting of 20 supergames each). The error probability ϵ_A is significantly smaller in the second part of all treatments compared to the first part. For ϵ_T, the decline between the two parts seems less steep, as perhaps one should expect for a more complex operation.

Looking at Table 11 we note that, if we consider separately the first and the second half of each treatment, the changes in strategies follow the model's predictions. In general, less harsh strategies become more frequent in the second part when errors become less frequent: AC, TfT and GT frequencies increase and AD frequency decreases. Table 11 shows that our data follow the predictions of the model, but also reveals additional complexity in the evolution and the structure of the errors. We argue that a view of their decline as exogenous processes of learning by doing will not qualitatively change the model's conclusions.

As we have shown earlier, in our data the error frequency decreases as subjects gain experience. The bottom panels of Figure 4 show this pattern in more detail. Both transition and action error probabilities decline over the session. Both types of error start from similar values in the two treatments and the patterns are not different across the two treatments. In the integrated treatment low IQ subjects converge closer to behavior of high IQ, suggesting that low IQ subjects learn from the high IQ subjects. If we consider the entire range of error probabilities that we observe in our data (see in particular figures O.3 and O.4 in the online appendix), we find an average change is in the range of 0 to 0.03.

Thus, we need to adjust the analysis for this situation, which is more complex than the simple case of constant error probabilities in our model. Looking at Figure 3, consider first the region in the simplex obtained as intersection of the basins of attraction for a strategy (for example that of Always Defect, AD), over the error probability in that range. Then do 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 AD), then the time path of frequencies will converge to AD even if the error probabilities 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 probabilities. In summary, the long run behavior is ultimately determined by the limit values of the error probabilities, which in our data are very low. However, starting with a higher error probability, the set of initial conditions on the fraction of strategies that eventually converge to AD will become larger, because of the initial movement in direction of AD.

We focus now on the evolution of strategies in our data. The top panels of Figure 4 report the frequency of strategies over time in the different treatments: high-IQ separated, low-IQ separated and integrated. As we said in several instances the data do not always allow to univocally identify the strategies. Therefore, to avoid counting strategies multiple times we make two additional assumptions:

1. Given our payoff structure and continuation probability, the Always Cooperate (AC) strategy is weakly dominated (see section 4.3), hence we assume that subjects are unlikely to choose this strategy and so rule it out every time this strategy is tied to others. Accordingly, Figure 4 presents the evolution of the share of subjects choosing AC, only when this strategy is uniquely identified.

2. To estimate the frequencies in each period of the remaining three strategies (i.e. AD, GT, TfT), we assume that they are played with the same probability when most likely and equally likely. Therefore, in periods where each of them is equally likely with at least another strategy, we estimate their respective frequency by dividing the total by the number of ties with other strategies. For example if TfT and GT are most likely and equally likely, we take the total frequency and divide by 2.

Comparing the two top panels of Figure 4 we note that the low-IQ tend to choose AD, more often when separated than when they are integrated with high-IQ subjects. The fraction of AD in the low-IQ separated sessions tends to stabilize just above a level of 0.20 in the long run, but initiates at a higher level (just above 0.40) in the first block of 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 initially high frequency of errors can bring the fraction of strategies during the early part of the session closer to high values of AD (see Figure 3), and, thus, in the basin of attraction of AD. The ensuing decline of error probabilities may not be able to reverse this initial drift towards the direction of Always Defect. The total frequency of Grim Trigger, GT, and Tit-for-Tat, TfT, converges to almost 100 per cent in the high-IQ separated and integrated sessions. The fraction of uniquely identified AC players is always small.

6 Conclusions

In spite of the many forces in society that tend to segregate individuals according to similar characteristics, a large part of social interaction occurs among individuals who are very different from one another in many respects. An important difference is that of their intelligence levels. This suggests that research on cooperation rates among individuals with different characteristics, in particular intelligence, may be illuminating and provide helpful guidance for policy.

It seems clear from several earlier experimental results, which have been replicated in the present paper, that higher cognitive skills tend to result in higher rates of cooperation. So we addressed here a natural question: What are the outcomes of strategic interactions among people of *heterogeneous* levels of intelligence? Our research offers three main results.

The first is that, in our experimental setting, cooperation rates in heterogeneous groups are closer to the higher cooperation rates that occur within groups composed exclusively by those of high intelligence, although the payoff for high-IQ players is lower than would be the case when playing only with their high-IQ peers. The aggregate payoff is higher when heterogeneous groups play together than when they play separately, but the interaction in heterogeneous pools is more advantageous to players of lower intelligence.

The second is that the higher cooperation rates of players of lower intelligence in integrated groups are generally due to the influence of the choices of the more intelligent players, who are more consistent in their implementation of strategies.

The third and final result concerns the observed shift towards harsher strategies of high intelligence players, when interacting with subjects of lower intelligence and higher error probabilities. We have argued that this is an instance of a more general phenomenon which can be explained by models of social learning.

The second and third main results are likely to be more general because they derive from the analysis of our model as well as by observing the experimental data.

To analyze the relation between error probabilities and distribution of strategies we propose an evolutionary game theory model where a population of players play a sequence of repeated games, and the frequency of strategies in the population is adjusted after every supergame. Studying the long run distribution of strategies and the size of their basins of attraction, we show that players choose more cooperative and forgiving strategies in environments where subjects commit few mistakes, and instead shift to harsher strategies when the error probabilities increase. The analysis of the data testing the model finds support for the main assumption, namely that error probabilities are negatively correlated with intelligence. It also provides evidence in favor of the main predictions of the model that the distribution of strategies depends on the error probability, with harsher strategies more frequent in populations with higher error probability.

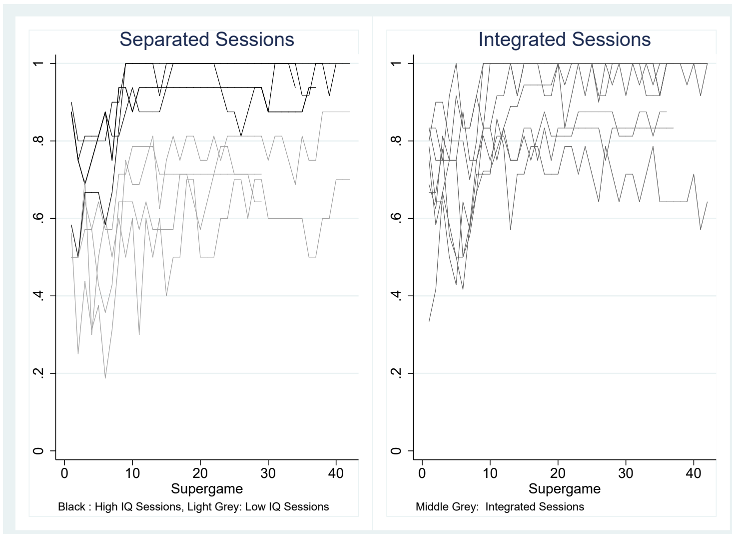
These results may provide useful guidance for policies affecting social interactions among groups that are heterogeneous in income or education levels, such as, for example, the interactions fostered by Moving to Opportunity (*MTO*) policies. In particular, we show that social interactions are likely to be mediated by differences in cognitive skills, which are likely to occur as consequences of these policies. Hence, our design can be considered a controlled *MTO* policy experiment, in which we can compare rates of cooperation and differences in strategies in groups separated according to cognitive skills with those rates that emerge in groups with different levels of cognitive skills.

Interesting extensions of these results should examine (in experiments and in theory) whether and how these results also emerge in more general games than Prisoner's Dilemma, and for more comprehensive sets of repeated game strategies. In repeated Prisoner's Dilemma games, the association between smaller error probabilities and less harsh strategies follows from the way in which the value function associating payoffs to strategies changes with the error probability. For example, Tit-for-Tat and Grim Trigger 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 Grim Trigger is higher than the loss for Tit-for-Tat; hence, the result is that the basin of attraction of Tit-for-Tat 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 in which players make mistakes, perhaps with a frequency associated with cognitive skill levels.

Further important insights could be gained by conducting a direct test of the best response evolutionary and rational learning models. This could be accomplished with an experimental design gathering appropriate measurement of the belief process. Such a study would require belief elicitation from participants during the session, either by direct observation (for example, by using eye-tracking measurements) or by surveys. Also, initial conditions on beliefs could be manipulated. This could be accomplished, for instance, by providing subjects at the beginning of the session with some information on the behavior of participants in a previous session, and checking with belief elicitation as to whether and how much this manipulation has been effective. This is a topic for current and future research.

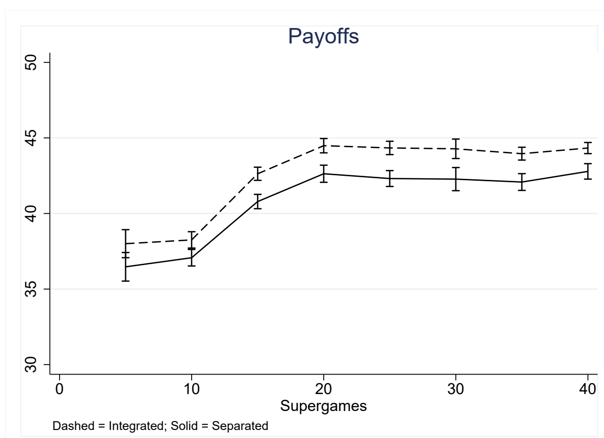
7 Figures and Tables

Figure 1: **Period 1 Cooperation for each supergame in all the Separated and Integrated sessions.** Average first period cooperation in each supergame and session. High- and low-IQ separated sessions are the black and grey lines respectively in the left panel. Integrated treatment sessions correspond to the middle grey lines in the right panel.



.

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



.

Figure 3: **Basin of attraction of A, G and T, with positive probability of 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 (top left panel) to 0.25 (bottom right panel). Payoff and discount factor are as in our experimental design.

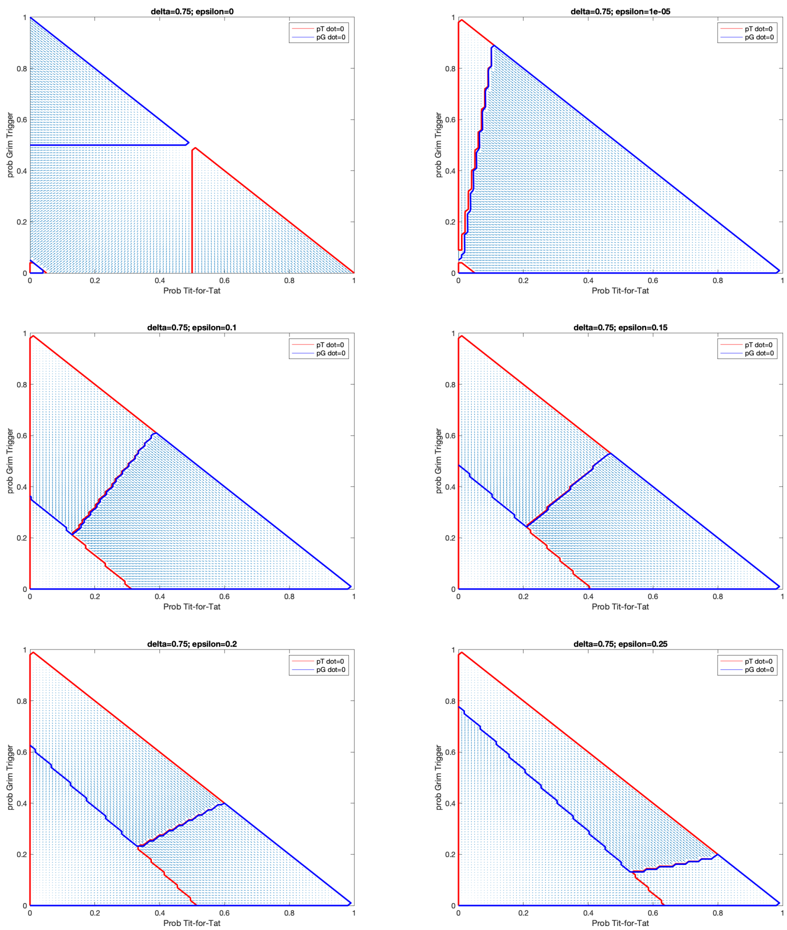


Figure 4: **Frequency of strategies and error probabilities over time.**

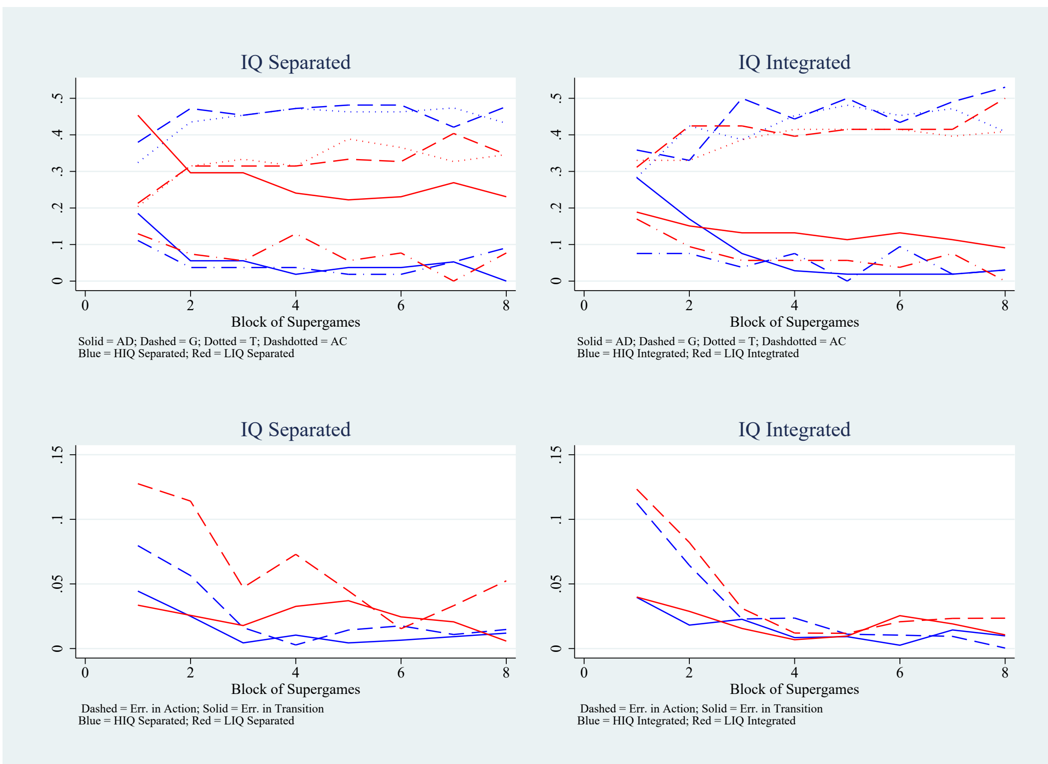


Table 3: **Prisoner's Dilemma.** C: Cooperate, D: Defect.

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

Table 4: **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 (OLS estimator). Integrated sessions are the baseline. Robust standard errors clustered at the session level in brackets; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$

	Supergame ≤ 20		Supergame > 20	
	Cooperate b/se	Payoff b/se	Cooperate b/se	Payoff 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 5: **Effects of IQ and other characteristics on the cooperative choice in first round 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	Round 1 Cooperate b/p
choice					
IQ	1.00889			1.00942	
	(0.6444)			(0.6396)	
High IQ Group		1.76893			0.59242
		(0.1401)			(0.1358)
Integrated Treatment			1.20968	1.17291	0.15483
			(0.6170)	(0.6737)	(0.6560)
Extraversion				0.87817	–0.09111
				(0.5544)	(0.6628)
Agreeableness				0.66879*	–0.39716*
				(0.0681)	(0.0851)
Conscientiousness				1.21574	0.20213
				(0.4599)	(0.4356)
Neuroticism				0.75337	–0.26813
				(0.3709)	(0.4035)
Openness				1.32202	0.27873
				(0.4504)	(0.4562)
Risk Aversion				0.79190***	–0.23160***
				(0.0063)	(0.0095)
Age				0.99517	–0.00198
				(0.9051)	(0.9605)
Female				1.04458	–0.00533
				(0.8941)	(0.9872)
Size Session				1.03245	0.02822
				(0.6398)	(0.6375)
N	214	214	214	214	214

Table 6: **Effects of separated 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 supergames. Integrated sessions are the baseline. Logit with individual fixed effects estimator. **Coefficients are expressed in odds ratios**. Note that in the second part of each session many subjects made the same choice throughout, and for this reason their observations needed to be excluded from the estimations of the model in columns 3 and 4 (this is also the reason why the estimated odds are very large). Similar regressions with random effects (which does not need variability of choices at the individual level avoiding this loss of observations) would deliver similar results. Standard errors in brackets; * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	Superg. ≤ 20		Superg. > 20	
	Cooperate b/se	Cooperate b/se	Cooperate b/se	Cooperate b/se
choice				
High IQ Session*Supergame	1.16022***	1.17081***	0.96562	1.02712
	(0.0582)	(0.0611)	(0.0643)	(0.0695)
Low IQ Session*Supergame	0.93705**	0.95825	1.09379**	1.11408**
	(0.0260)	(0.0273)	(0.0468)	(0.0500)
Supergame	1.13538***	1.09608***	0.99093	0.93260*
	(0.0283)	(0.0281)	(0.0296)	(0.0341)
1st Per. Part. Coop. Rates until s-1		33.09363***		168284.00972**
		(15.4975)		(1.01e+06)
Partner Coop Rates until t-1		0.77164		19513.03094**
		(0.2548)		(9.50e+04)
Average Supergame Length	2.00252***	2.23329***	5.70321**	8.25299**
	(0.2402)	(0.2921)	(4.5775)	(6.9723)
N	2280	2280	654	654

Table 7: **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 block of 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 Separated b/p	High IQ Separated b/p	Low IQ Integrated b/p	High IQ Integrated 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 Separated b/p	High IQ Separated b/p	Low IQ Integrated b/p	High IQ Integrated 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 8: **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 $t-1$, that is, $(C,C)_{t-1}$. Integrated is a dummy for the integrated treatment. Probit with correlated random effect estimator: the variables average cooperation in 1^{st} periods and cooperation in the 1^{st} round are included in the regression but not reported. Robust standard errors clustered at the session level 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		
Integrated*$(C,C)_{t-1}$	0.15679	–0.03105
	(0.1888)	(0.1134)
Integrated*$(D,D)_{t-1}$	–0.32689*	0.01411
	(0.1807)	(0.2423)
Integrated*$(D,C)_{t-1}$	–0.15450	–0.18132
	(0.1554)	(0.1626)
Integrated*$(C,D)_{t-1}$	–0.28643**	0.06988
	(0.1283)	(0.1873)
$(D,D)_{t-1}$	–3.43989***	–3.40710***
	(0.1919)	(0.1714)
$(D,C)_{t-1}$	–2.37170***	–2.74250***
	(0.2053)	(0.0851)
$(C,D)_{t-1}$	–2.67842***	–2.79165***
	(0.0886)	(0.1517)
N	10343	10003

Table 9: **Determinants of error probabilities across individuals.** Dependent variables are the error probabilities. The variable Integrated is a dummy for the integrated treatment. GLS estimator with random-effects. Errors are clustered at the session level; * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	Err. in Action			Err. in Transition		
	Prob. b/se	Prob. b/se	Prob. b/se	Prob. b/se	Prob. b/se	Prob. b/se
IQ	–0.0011*	–0.0010	–0.0015**	–0.0005**	–0.0005**	–0.0007***
	(0.0007)	(0.0007)	(0.0007)	(0.0002)	(0.0002)	(0.0002)
Integrated			–0.1025			–0.0439
			(0.1355)			(0.0344)
IQ * Integrated			0.0009			0.0004
			(0.0013)			(0.0003)
Block of Superg.		–0.0128***	–0.0127***		–0.0031***	–0.0031***
		(0.0012)	(0.0012)		(0.0008)	(0.0008)
N	1540	1540	1540	1540	1540	1540

Table 10: **Individual strategies and errors in the different treatments.** The table reports the share of the Maximum Likelihood (ML) Strategies and the error probabilities for each treatment. They are estimated using the algorithm illustrated in section E of the appendix. Given that some ML strategies are not unique (i.e. their likelihood is identical to at least another strategy) their sums will not add up to one. The asterisks indicates whether the differences are significantly different from 0. * $p-value<0.1$, ** $p-value<0.05$, *** $p-value<0.01$.

Treatments	**IQ Separated**			**IQ Integrated**				**Separated vs. Integrated**			
IQ Group	*High IQ*	*Low IQ*		*All*	*High IQ*	*Low IQ*					
	1	2	Diff. 1 - 2	3	4	5	Diff. 4 - 5	Diff. 1 - 3	Diff. 3 - 2	Diff. 1 - 4	Diff 2 - 5
Strategies											
AC	0.664	0.230	0.434***	0.542	0.564	0.520	0.044	0.122***	0.312***	0.100***	-0.290***
TfT	0.802	0.520	0.282***	0.729	0.755	0.703	0.052*	0.073***	0.209***	0.047	-0.183***
GT	0.818	0.511	0.307***	0.751	0.777	0.725	0.052*	0.067**	0.240***	0.041	-0.214***
AD	0.060	0.290	-0.230***	0.109	0.084	0.134	-0.05**	-0.049***	-0.181***	-0.024	0.156***
Probability of Error											
Error in Action (ϵ_A)	0.028	0.071	-0.043***	0.038	0.033	0.042	-0.009	-0.010*	-0.033***	-0.005	0.029***
Error in Transition (ϵ_T)	0.015	0.027	-0.012**	0.018	0.016	0.020	-0.004	-0.003	-0.009**	-0.001	0.007

Table 11: **Individual strategies and errors in the different treatments and periods** The table reports the share of the Maximum Likelihood (ML) Strategies and the error probabilities for each treatment separately for the first and the second halves of each treatment. They are estimated using the algorithm illustrated in section E of the appendix. Given that some ML strategies are not unique (i.e. their likelihood is identical to at least another strategy) their sums will not add up to one. The asterisks indicates whether the differences are significantly different from 0. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

Treatments	**IQ Separated**						**IQ Integrated**								
IQ Group	*High IQ*			*Low IQ*			*All*			*High IQ*			*Low IQ*		
Block of Superg.	1 to 4	5 to 8	Diff.	1 to 4	5 to 8	Diff.	1 to 4	5 to 8	Diff.	1 to 4	5 to 8	Diff.	1 to 4	5 to 8	Diff.
Strategies															
AC	0.574	0.780	0.206***	0.204	0.273	0.069	0.450	0.643	0.193***	0.458	0.682	0.225***	0.443	0.604	0.161***
TfT	0.741	0.881	0.140***	0.454	0.629	0.175***	0.651	0.815	0.164***	0.665	0.854	0.189***	0.637	0.776	0.139***
GT	0.764	0.887	0.123***	0.449	0.614	0.165***	0.672	0.839	0.166***	0.684	0.880	0.196***	0.660	0.797	0.136***
AD	0.079	0.036	-0.043*	0.324	0.235	-0.089*	0.146	0.068	-0.079***	0.142	0.021	-0.121***	0.151	0.115	-0.036
Prob. of Error															
Error in Action (ϵ_A)	0.039	0.015	-0.024***	0.090	0.038	-0.052***	0.059	0.015	-0.044***	0.056	0.009	-0.047***	0.062	0.019	-0.043***
Error in Transition (ϵ_T)	0.021	0.007	-0.014**	0.027	0.025	-0.002	0.022	0.013	-0.010**	0.022	0.009	-0.013**	0.023	0.017	-0.006

References

Alaoui, L., Janezic, K. A. and Penta, A. (2020). Reasoning about others' reasoning, *Journal of Economic Theory* **189**: 105091.

Alaoui, L. and Penta, A. (2016). Endogenous depth of reasoning, *The Review of Economic Studies* **83**(4): 1297–1333.

Alaoui, L. and Penta, A. (2018). Cost-benefit analysis in reasoning, *Technical report*, Barcelona GSE, number 1062.

Aoyagi, M., Bhaskar, V. and Fréchette, G. R. (2019). The impact of monitoring in infinitely repeated games: Perfect, public, and private, *American Economic Journal: Microeconomics* **11**(1): 1–43.

Aoyagi, M. and Fréchette, G. (2009). Collusion as public monitoring becomes noisy: Experimental evidence, *Journal of Economic Theory* **144**(3): 1135–1165.

Aubin, J.-P. and Cellina, A. (1984). *Differential Inclusions. Set-Valued Maps and Viability Theory*, Springer Verlag, Grundlehren der mathematischen Wissenschaften.

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.

Boyd, R. (1989). Mistakes allow evolutionary stability in the repeated prisoner's dilemma game., *Journal of theoretical biology* **136 1**: 47–56.

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.

Jagelka, T. (2019). Are economists' preferences psychologists' personality traits?, *Technical report*, University of Bonn.

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. (2010). *Population games and evolutionary dynamics*, MIT press.

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.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* **4**: 25–55. 10.1007/BF01766400.
URL: *http://dx.doi.org/10.1007/BF01766400*

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

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*, MIT press.

A Evolution of Strategies with Errors

In this section we study in detail the evolution of strategies with errors, in the general model both with errors in action and in transition. Our ultimate aim is to present, in the final section A.4, a model of learning in games that is close to the experimental environment we adopt. The section will first consider the simple case in which there are no errors (section A.1) to establish a benchmark. This will be used later to demonstrate that the model with errors, even of arbitrarily small size, is qualitatively different from the no-error case. In section A.2 we give a complete analysis of the model described verbally in the main text, where errors in transition only are considered. In the following section A.3 we extend the analysis to both types of error. Finally in section A.4 we consider a rational learning model. We estimate the learning model in the data, providing a link between the evolutionary models and the data analysis. As in section 4.3, to lighten notation we shorten the names of the strategies as A, G, T, where A stands for *Always Defect* strategy, G for *Grim Trigger*, and T for *Tit-for-Tat*.

A.1 No Errors

We begin our analysis with the benchmark case of no errors; in our notation, we require $\epsilon_T = \epsilon_A = 0$. In this case, the payoff matrix for the strategy choice game is the one reported earlier in table 2 of the main text. This game is special, because the two "actions" G and T are interchangeable. Thus, the analysis of the game with three actions can be reduced to the analysis of the *reduced game* obtained by collapsing G and T into a single action, denoted SC (sophisticated cooperation), and producing a game with two actions $\{A, SC\}$, and payoff matrix:

Table A.1

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

We denote by μ_R the strategies in the reduced game, and introduce an important threshold value for δ:

$$\overline{\delta} \equiv \frac{t-c}{t-d}. \tag{A-1}$$

.

Proposition A.1. *When $\delta > \overline{\delta}$ the reduced game has two equilibria in pure strategies (namely (A, A) and (SC, SC)) and a mixed strategy equilibrium denoted $\mu_R(\cdot, \delta)^*$. For any such equilibrium there is a corresponding continuum of equilibria in the strategy choice game, where each equilibrium is obtained by assigning the probability $\mu_R(GT, \delta)^*$ arbitrarily to the strategies G and T.*

Proof. By the assumption stated in equation (2), $d > s$ and so $d > (1-\delta)s + \delta d$. When $c < (1-\delta)t + \delta d$, A is a dominant strategy in the reduced game, 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 instead $c > (1-\delta)t + \delta d$, that is when $\delta > \overline{\delta}$. In this case there are three equilibria, with the mixed strategy equilibrium assigning a probability to A for the given δ given by:

$$\mu_R(A, \delta)^* \equiv \mu^* = \frac{c - (1-\delta)t - \delta d}{c - (1-\delta)t - \delta d + (1-\delta)d} \tag{A-2}$$

□

Corresponding to these equilibria, there is a set of steady states in the evolutionary dynamic. The dynamic under proportional imitation has a natural long run behavior: when the initial 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 initial conditions alters the long run behavior. More precisely:

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

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

$$\{(\mu^*, (1 - \mu^*)p, (1 - \mu^*)(1 - p)) : p \in [0, 1]\} \tag{A-3}$$

and the edge

$$\{(0, p, (1 - p)) : p \in [0, 1]\}; \tag{A-4}$$

2. *the paths 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 (A-4) above;*

3. *the steady states in the set (A-3) 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*:*

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

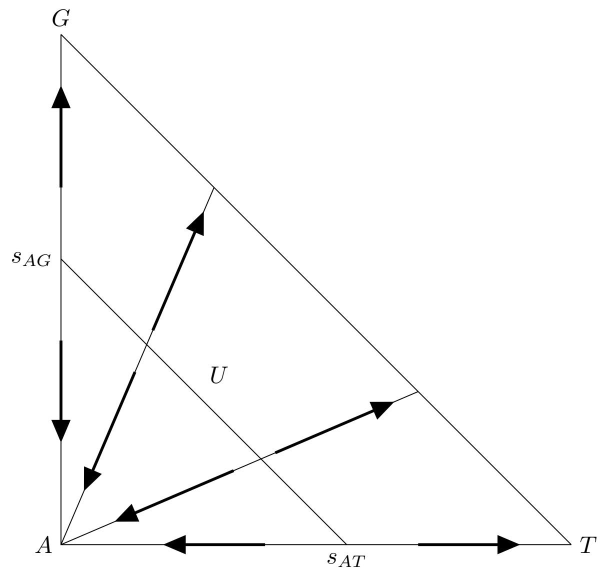
Thus a path with initial condition $\mu(0, A) < \mu^*$ *converges to the steady state given by:*

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

In conclusion, evolutionary dynamics in the case we are considering (automata with no error) cannot select among possible relative frequencies of G and T, and thus cannot address the issue of whether a strategy such as T is more or less frequent in the long run than the harsher 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 A.1 illustrates the dynamic behavior of the proportional imitation model when no errors occur. 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 with points $(\mu(G), \mu(T))$, 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.

The dynamic behavior with Best Response dynamics is very similar (see the first panel of Figure 3), but in such a degenerate case (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.

Figure A.1: **Proportional Imitation dynamics with no error.** All points on the segment joining s_{AG} and s_{AT} (denoted by U in the figure) 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 of these lines are illustrated here.



A.2 Errors in Transition

We now introduce in our benchmark model the possibility of errors in transition, allowing $\epsilon_T > 0$.

We first have to define how errors affect the payoff matrix introduced in table 2 for the case of no errors. To compute the payoff from the choice of the strategy when transition errors are possible we need to extend the state space to explicitly include the automata that are produced by errors, distinguishing two automata on the basis of the initial internal state. So we introduce automata G_c and G_d, defined as having the same transition and same action choice function as G, but having c and d respectively as initial state, where c is the state in which cooperation is chosen. So G_d is different from A because the state of the automaton G_d may transit back to c (although this can only happen by mistake), whereas the state of A can never transit to a state in which the action C is chosen. With no errors, G_d and A are the same because for any action profile the transition in G_d from the defect state is back to the defect state. With errors in transition, there may by mistake a transition from the initial defect state into the cooperate state. T_c and T_d are defined similarly.

We then introduce an extended strategy set S (to distinguish it from the set M defined in equation (4)) which includes these two new automata:

$$S \equiv \{A, G_c, G_d, T_c, T_d\} \tag{A-5}$$

where G_c is the G (Grim Trigger) automaton with c as initial state, (d initial state for G_d). There is a unique action determined by an automaton $s \in S$, denoted by $a(s)$. The payoff from a pair of choices of initial automata made by the two players is determined by a simple recursive equation on functions defined on the product space Ω:

$$\Omega \equiv S \times S \tag{A-6}$$

with generic element $\omega = (s^1, s^2)$. With this notation we can write the transition of 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 (still denoted by P) with $P(s'; s, a^1, a^2)$ describing the probability of transiting to s' if the current state is s and the action profile is (a^1, a^2).[32]

We now turn to the definition of the transition function with errors. We let $\mathcal{Q}$ be the set of stochastic matrices on Ω, that is $Q \in \mathcal{Q}$ if $Q : \Omega \to \Delta(\Omega)$. First, we let the transition with no errors to be denoted by Q_0, where

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

$$Q_0(\omega'; \omega) \equiv P(r^1; s^1, a(\omega)) P(r^2; s^2, a(\omega)) \tag{A-7}$$

We denote by $q^i(\omega)$ the state in S for player i to which player i transits given the current pair ω, taking into account the corresponding action choice. Finally, we let $Q_{\epsilon_T} \in \mathcal{Q}$ at error rate ϵ_T to be:

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

The stage game payoff function u was introduced in section 4.3, immediately following table 1. We can now define a payoff function $\Omega \to \mathbb{R}$, still denoted by u, defined at $\omega = (s^1, s^2)$ by:

$$u(\omega) = u(a^1(s^1), a^2(s^2)); \tag{A-8}$$

A.2.1 Value Function

Players in the strategy choice game choose an element in the set $\{A, G, T\}$, but due to the presence of errors the value function must be 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 a value function $V : \Omega \to \mathbb{R}$. We write $V(\omega; \epsilon_T, \delta)$ when we want to emphasize that the value depends on the parameters.

Lemma A.3. *For fixed u, the function $(\epsilon_T, \delta) \longrightarrow V(\cdot; \epsilon_T, \delta)$ is analytic, and hence continuous and differentiable.*

Proof. The function V is the unique solution of the functional equation with the two parameters

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

(u, Q_{ϵ_T}):

$$V = (1-\delta)u + \delta Q_{\epsilon_T} V \tag{A-9}$$

Since $\delta < 1$, the inverse matrix $(I - \delta Q_{\epsilon_T})^{-1}$ exists and therefore:

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

The derivative of V with respect to the error parameter is

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

□

The analysis of the function V is considerably simplified if we observe that Ω is partitioned into invariant sets under the transition Q_{ϵ_T}. We say that a subset Ω_i of Ω is invariant under Q_{ϵ_T} if and only if:

$$\forall \omega \in \Omega_i, Q_{\epsilon_T}(\Omega_i;\omega) = 1.$$

For example the set $\Omega_{AG} \equiv \{(A, G_c), (A, G_d)\}$ is invariant, because the automaton of the first player can only be A, and the automaton of the second player can only transit to G_c or G_d. The other eight sets are denoted in a similar, natural, way. Overall we have the following partition of Ω:

$$\mathcal{P}(\Omega) \equiv \{\Omega_{AG}, \Omega_{AA}, \Omega_{AT}, \Omega_{GA}, \Omega_{GG}, \Omega_{GT}, \Omega_{TA}, \Omega_{TG}, \Omega_{TT}\}. \tag{A-10}$$

each of which is invariant. 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 (A-9) with (u, Q_{ϵ_T}) replaced by $(u_i, Q_{\epsilon_T,i})$. These equations can be solved and analyzed independently.

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

Lemma A.4. *For all $\epsilon_T > 0$, (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 in table A.2.

Table A.2

	G_c, A	G_d, A
G_c, A	ϵ_T	$1-\epsilon_T$
G_d, A	ϵ_T	$1-\epsilon_T$

Using equation (A-9), we can solve for $V(G_c, A)$; after some algebraic manipulations we find:

$$\begin{aligned} V(G_c, A) &= (1-\delta(1-\epsilon_T))u(G_c, A) + \delta(1-\epsilon_T)u(G_d, A) \\ &= (1-\delta(1-\epsilon_T))s + \delta(1-\epsilon_T)d \end{aligned}$$

Therefore for all $\epsilon_T > 0$,

$$V(G_c, A) < d = V(A, A). \tag{A-11}$$

The transition matrix restricted to Ω_{AT} is the same as the one we reported in table A.2 for Ω_{AG}, and so we similarly conclude:

$$V(T_c, A) < V(A, A). \tag{A-12}$$

□

Lemma A.5. *The value function equation can be decomposed into nine independent equations, one for each of the invariant sets of the set Ω. Also the following equations hold:*

1. $V(A, A) = d$
2. $V(A, G) = V(A, T) = t(1-\delta(1-\epsilon_T)) + d\delta(1-\epsilon_T)$
3. $V(G, A) = V(T, A) = s(1-\delta(1-\epsilon_T)) + d\delta(1-\epsilon_T)$

Proof. The equation for the value of $V(A, A)$ follows from the fact that the singleton $\{(A, A)\}$ is invariant. The equations for the values of $V(A, G)$ and $V(A, T)$ follow from the computations in the proof of lemma (A.4). □

Uniform Errors Using the invariant sets allows us to compute the payoff on the case of uniform error, that is when $\epsilon_T = 1/2$. This case gives us a boundary condition for the study of the dynamic behavior. In particular the analysis will tell us when it is impossible that (G, G) is an equilibrium profile in the strategy choice game.[33]

Lemma A.6. *When $\epsilon_T = 1/2$ 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 we have denoted:

$$m_s \equiv (1-\delta)s + \delta\frac{s+d}{2}, m_t \equiv (1-\delta)t + \delta\frac{t+d}{2}, \tag{A-13}$$

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

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

[33] 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. In the alternative formulation where an automaton is chosen, and an error in the choice of the initial state is possible, the conclusion of lemma (A.6) is even easier to prove.

Proof. For any element $i \in \mathcal{P}(\Omega)$ (the latter set is defined in (A-10), so i is a subset of Ω), we have that the transition restricted to i is:

$$Q_{(1/2),i} = \frac{1}{\#i} U_i \equiv M_i \tag{A-14}$$

where U_i is a square matrix of 1's of dimension $\#i$ (the cardinality of the set i). Note:

$$M_i^2 = M_i. \tag{A-15}$$

The value function equation restricted to states in i is:

$$V_i = (1-\delta)u_i + \delta M_i V_i \tag{A-16}$$

Given equation (A-15), equation (A-16) implies

$$M_i V_i = (1-\delta) M_i u_i + \delta M_i V_i$$

which in turn implies, since $1 - \delta > 0$:

$$M_i V_i = M_i u_i \tag{A-17}$$

Now (A-16) and (A-17) give the formula for the value in terms of the payoffs:

$$V_i = (1-\delta)u_i + \delta M_i u_i \tag{A-18}$$

The rest follows from simple algebra.

For the last statement, since $t > c$ and $d > s$ (from our assumption in equation (2) on the stage game payoffs) we conclude that $m_t > m_c$. From $d > s$ we get that $d > m_s$ holds for any δ. Thus A is a dominant action in the strategy choice game. □

Small Errors We now know that for "large" errors (that is, when $\epsilon_T = 1/2$) the strategy choice game has a unique equilibrium, namely (A, A). The dynamic behavior for small ϵ_T on the sides $\mu(T) = 0$ and $\mu(G) = 0$ is similar to that of the no error model, with a unique steady state that changes continuously in ϵ_T around $\epsilon_T = 0$. 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 forgiving strategy T and the harsher strategy G.

In our analysis, "small ϵ_T" means smaller than a critical value that we introduce now. Note that this critical value might not be numerically small. We let:

$$\overline{\epsilon_T} \equiv \sup\{\epsilon_T : \min\{V(G,G), V(T,G)\} > V(A,G) \tag{A-19}$$

$$\text{and } \min\{V(T,T), V(G,T)\} > V(A,T)\}$$

Lemma A.7. *For any payoff of the stage game defined in section 4.3 that satisfies the condition in equation (2), $\overline{\epsilon_T} > 0$. If it is finite, then the value is achieved. There are payoffs satisfying the standing assumption on stage game payoffs stated in equation (2) for which $\overline{\epsilon_T}$ is finite.*

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

$$\begin{aligned} V(G,G) &= V(T,G) \\ &= c \\ &> t(1-\delta) + d\delta \\ &= V(A,G) \end{aligned}$$

where the strict inequality follows from our assumption in equation (2). Similarly at $\overline{\epsilon_T} = 0$, $V(T,T) = V(G,T) > V(A,T)$. Thus we conclude that $\overline{\epsilon_T} > 0$ from lemma (A.3). The second claim also follows from lemma (A.3). The last claim follows from lemma (A.6), which gives as simple sufficient condition for $\overline{\epsilon_T}$ to be finite. ☐

In particular $\overline{\epsilon_T}$ is finite for the stage game with payoffs as in our experimental design. The next lemma describes the dynamic for small ϵ_T in PID. Equation (5) makes clear that the edges of the simplex are invariant under PID; if an initial condition has the probability of G equal to 0, the entire path from that initial condition has the probability of G equal to 0.

Lemma A.8. *For the proportional imitation model, and for* $0 < \epsilon_T < \overline{\epsilon_T}$*:*

1. *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* $\{A, G\}$*);*
2. *there is a unique steady state on the side* $\mu(A) = 0$*, and this is determined in the limit* $\epsilon_T \to 0$ *by the ratio:*
$$\mu(G)^* = \frac{V_{\epsilon_T}(T,T) - V_{\epsilon_T}(G,T)}{V_{\epsilon_T}(T,T) - V_{\epsilon_T}(G,T) + V_{\epsilon_T}(G,G) - V_{\epsilon_T}(T,G)} \tag{A-20}$$
where V_{ϵ_T} *is the derivative with respect to* ϵ_T*.*

Proof. Point (1) follows from the equation describing the dynamics. For point (2), note first that for $\epsilon_T > 0$ then the steady state is determined by a frequency of the strategy G, $\mu(G)^*$ such that

$$\mu(G)^* = \frac{V(T,T) - V(G,T)}{V(T,T) - V(G,T) + V(G,G) - V(T,G)} \tag{A-21}$$

As $\epsilon_T \to 0$ both numerator and denominator in the right hand side of (A-21) 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 other four steady states are unstable.

We now turn to the main question we posed, namely the relationship between the probability of error (modeled by the ϵ_T parameter) and the frequency of strict and harsh strategies, with the strict harshness order:

$$A \succ G \succ T \tag{A-22}$$

For any evolutionary dynamics, we denote the steady states (when they exist) which are located on the edges of the simplex as $s_{A,T}$, $s_{A,G}$ and $s_{G,T}$ respectively, where the subscript indicates the extreme points of the edge. For example, in the case of the PID, $s_{A,T}$ is the point in the simplex of the form $(p, 0, 1-p)$ with $p \in (0,1)$ at which the time derivative in equation (5) is equal to 0. As we indicate in section A.2.3 in our discussion and figures we consider the projection of the simplex of the coordinates indicating the probability of G and T: we do not distinguish in the notation between the steady state as element of the simplex and its projection on the plane. For illustration of $s_{A,T}$, $s_{A,G}$ and $s_{G,T}$ we refer to Figures A.1 and O.7.

A.2.2 Proportional Imitation Dynamics

In the following analysis we will examine the size of the basin of attraction of a strategy under PID, using the location of the three points $s_{A,T}$, $s_{A,G}$ and $s_{G,T}$ to provide orientation.

As we noted earlier, the edges of the simplex are invariant under PID. Let us now consider the dynamic of the process restricted to the edges. It is clear that this evolution is determined only by the value function restricted to the product of the two strategies at the vertices of the edge; so we can focus on the corresponding two actions game.

Consider now 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 the probability of action a given by $p^* = \frac{g(b)}{g(a)+g(b)}$. If we consider the PID in this simple two-action game, the basin of attraction of the action a is the set $\{(p, 1-p) : 1 \geq p > p^*\}$, a sub-interval of the unit interval of length

$$1 - p^* = \frac{g(a)}{g(a) + g(b)} \tag{A-23}$$

which we call the *size* of the basin of attraction of A. As intuition may suggest, the size of the basin of attraction of strategy a is increasing in the ratio of the gain of action a over that of action b.

We apply this analysis to our three actions game.[34] To illustrate consider the edge of the simplex between the vertices A and T, with steady state within the edge (if it exists) $s_{A,T}$. The size of the basin of attraction of A (the harshest of the two strategies) within the edge will be denoted $sba_{A,T}$, and is equal to $1 - s_{A,T}(A)$. Similarly we denote $sba_{A,G} \equiv 1 - s_{A,G}(A)$ and $sba_{G,T} \equiv 1 - s_{G,T}(G)$, taking in each case the harshest of each pair of strategies (A and G respectively) as reference. When we want to emphasize the dependence of the steady state on parameters, we write $sba_{A,T}(\epsilon_T, \delta, u)$ and so on.

An increase in the values of $sba_{A,T}$, $sba_{A,G}$ 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 $sba_{G,T}$ increases that of G at the expense of T. The following proposition summarizes what we know for the two extreme values of $\epsilon_T = 0$ and $\epsilon_T = 1/2$:

Proposition A.9. *Under proportional imitation,*

$$sba_{A,G}(0, \delta, u) = sba_{A,T}(0, \delta, u) = \frac{(1-\delta)(d-s)}{c - (1-\delta)t - (1-\delta)s - \delta d + (1-\delta)d} \tag{A-24}$$

$$sba_{A,G}(1/2, \delta, u) = sba_{A,T}(1/2, \delta, u) = sba_{G,T}(1/2, \delta, u) = 1 \tag{A-25}$$

Proof. Use the values in table A.1 and apply (A-23). □

[34] Figures A.1 or O.7 may provide a useful illustration.

At $\epsilon_T = 0$ the basin of A is the triangular region below the straight-line segment joining $s_{A,G}$ and $s_{A,T}$. As ϵ_T reaches the value 1/2 any point in the simplex (except the side between G and T) converges to A over time. The strategy A may become the only strategy surviving for values of the error probability smaller than 1/2, as we next demonstrate.

We analyze the basin of attraction for the specific numerical values of the payoffs used in the experiment. In section C of the online 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, we can use the lemmas (A.3) and (A.5), to compute the value function and analyze equilibria and dynamic behavior.

Figure A.2 uses the simple formula for the size of the basin of attraction. The figure reports the three values $sba_{A,G}$, $sba_{A,T}$, $sba_{G,T}$ as indicated in the legend: these are the three lines within each panel. The comparison is made for different values of ϵ_T (moving along the x-axis in each panel) and δ (along the three panels). The value of ϵ_T at which the line reaches the value 1 is the value at which the steady state giving positive probability to the less harsh strategy disappears, and the entire interval converges to the harsher strategy. For instance, the value of ϵ_T at which the line reporting the size of the basin for A within the pair $\{A, T\}$ is the value at which $s_{A,T}$ disappears, $sba_{A,T}$ becomes equal to 1, and the entire interval converges to A.

It is clear from the figure that the steady state $s_{A,T}$ disappears last for all the values of δ considered. Instead the order of disappearance of $s_{A,G}$ and $s_{G,T}$ switches: $s_{G,T}$ before $s_{A,G}$ for low δ, $s_{G,T}$ after $s_{A,G}$ for higher δ. We denote as $\hat{\delta}$ the value of δ at which $s_{G,T}$ and $s_{A,T}$ disappear at the same value of ϵ_T; with the payoffs used in the experimental design, the value of $\hat{\delta}$ is 0.82.

More in detail, the initial conclusions we draw from Figure A.2 is:

Conclusion A.10.

1. *The values $sba_{A,G}$ and $sba_{A,T}$ are monotonically increasing in ϵ_T.*
2. *For any value of the continuation probability, as the error probability becomes larger, the basin of attraction of less harsh strategies (G and T) vanishes. The relative size of the basin of attraction of T declines compared to that of G as the error probability increases.*
3. *With sufficiently high continuation probability and small error, the basin of attraction of A is arbitrarily small.*

The intermediate value $\hat{\delta}$ marks the boundary of an interesting division of the set of δ's into two sub-regions. We define precisely $\hat{\delta}$ and its significance, focusing on the interior of the projection of the simplex:

Conclusion A.11.

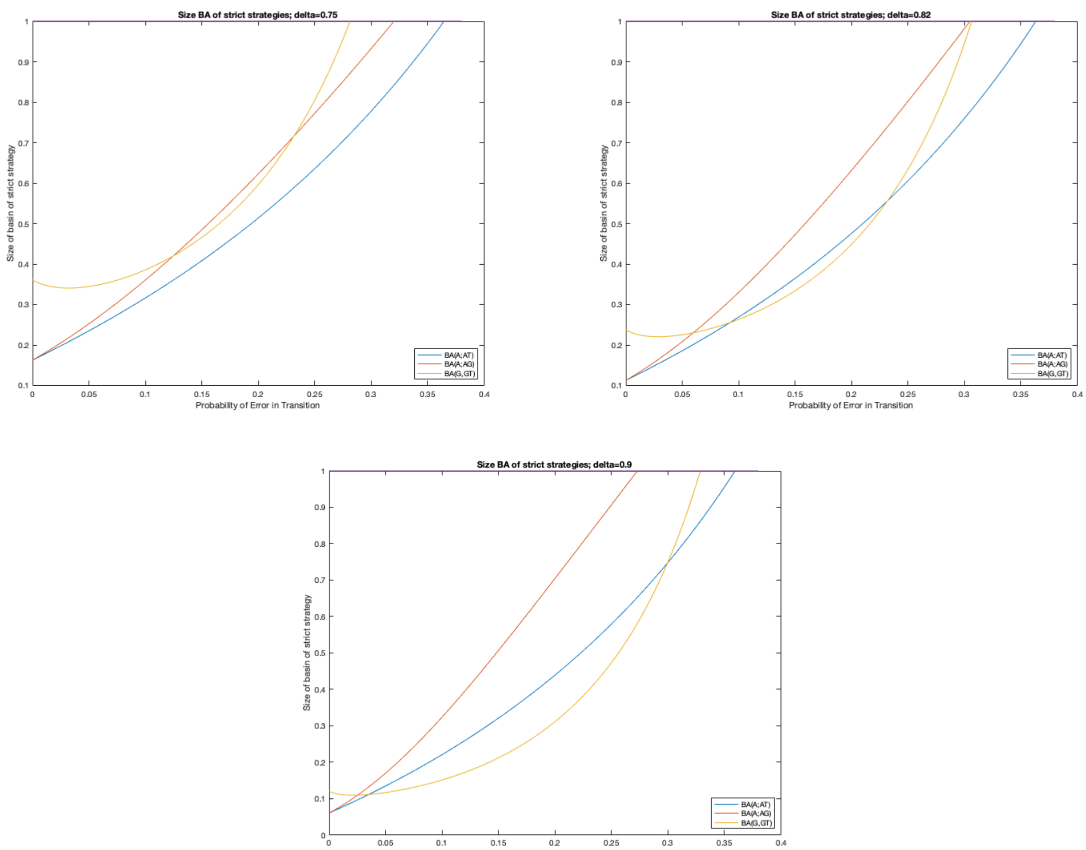
1. *As ϵ_T increases from 0 to 1/2, the interior of the simplex is initially partitioned into three regions (with attractors A, G and T); then into two regions with two attractors, and finally only into a single region (with attractor A);*
2. *If $\delta > \hat{\delta}$ - in the intermediate region of ϵ_T - the two attractors are A and T; if $\delta < \hat{\delta}$ instead the two attractors are A and G.*

For lower values ($\delta < \hat{\delta}$) the basin of attraction of T disappears entirely as s_{GT} collapses into the vertex T, and the entire edge between the vertices G and T of the simplex converges to G. With lower continuation probability ($\delta < \hat{\delta}$) but high probability of errors the strategy T does not survive. For higher values both G and T survive, but both lose frequency at the expense of A. In conclusion:

Conclusion A.12. *Strategy T can survive only with low ϵ_T or with high δ.*

With small errors, an appropriately modified Poincare-Hopf index can be calculated. The index of the three stable pure strategies is 1 in all cases; the others (the two mixed strategies on the edge between A and T and the edge between A and 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.

Figure A.2: **Size of basin of attraction of strategies A and G on the sides of the simplex.** Top to bottom panel: values of δ equal to 0.75, 0.82, 0.9 respectively.



A.2.3 Best Response Dynamics

The best response dynamic allows us to identify clearly the basin of attraction of the three strategies by simple inspection of the phase portrait. To understand the difference between this case and the proportional imitation dynamics, it is useful to keep in mind that there may be steady states for the PID on the boundary of the simplex that are not Nash equilibria of the entire game; in this case a steady state for PID may fail to be a steady state of the BRD.

To be precise, let $NE(\{A, G, T\})$ be the set of Nash equilibria of the strategy choice game where the player can choose any strategy in $\{A, G, T\}$. We will refer to this as the unrestricted game. For any two-strategy subset $\{r, s\}$ of the set $\{A, G, T\}$, let $N(\{r, s\})$ be the Nash equilibria of the reduced game, where players can only choose from $\{r, s\}$. It may occur that a steady state of the PID at the boundary is not a Nash equilibrium of the unrestricted game, 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 $\epsilon_T = 0.35$ and $\delta = 0.9$ [35] the strategy T is weakly dominated by G and $V(G, T) > V(T, T)$. Thus, s_{AT} (which is a mixed strategy of the unrestricted game, assigning 0 to G) is not part of a Nash equilibrium of the unrestricted game, so it is not a steady state of BRD but it is a steady state of the PID. The same may be true for pure strategies: for example, the strategy G is a steady state in the game induced by $\epsilon_T = 0.3$ and $\delta = 0.9$ in Table O.19 of the online appendix, the pure strategy G is a steady state of the PID, but is not a Nash equilibrium of the complete game, because $V(A, G) > V(G, G)$.

In the interior of the projection of the simplex the situation is considerably simpler:

Proposition A.13. *With best response dynamics,*

1. *for $\epsilon_T < \overline{\epsilon_T}$, all three strategies survive in the long run (that is, they have a basin of attraction with non-empty interior); 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 (A.10), (A.11) and (A.12) hold in the case of BRD as well; in particular, for $\epsilon_T > \overline{\epsilon_T}$ only two strategies survive in the long run: they are $\{A, G\}$ for $\delta < \hat{\hat{\delta}}$, and $\{A, T\}$ for $\delta > \hat{\hat{\delta}}$.*

The lines marking the boundary of the basins of attraction in the BRD and PID may be different (for example, those in BRD are straight lines, while those in PID are not) but the end points are common, hence qualitatively the basins are similar. In particular the analysis in the previous section identifying different transitions for values of δ above and below $\hat{\hat{\delta}}$, as ϵ_T changes, holds.

A.3 Errors in Transition and in Action

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 previous sections is the introduction of the error in action transition matrix. Recall that ϵ_A is the probability of an error in the choice of the action at a state in the automaton, and $A \equiv A^1 \times A^2$. Let $Pr(a; \omega, \epsilon_A)$ 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 ϵ_A. The action choice with errors at a state is a stochastic matrix $A_{\epsilon_A} : \Omega \to \Delta(\Omega \times A)$ defined by

$$A_{\epsilon_A}(\omega)(\omega', a) \equiv \delta_\omega(\omega') Pr(a; \omega, \epsilon_A) \tag{A-26}$$

[35] The precise values are reported in the third panel of Table O.19 of the online appendix.

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 $Q_{\epsilon_T,\epsilon_A}$ in equation (A-27). To illustrate the definition of A_{ϵ_A} consider for example:

$$A_{\epsilon_A}(G_c, G_d)(G_c, G_d, C^1, D^1) = (1 - \epsilon_A)^2$$

The transition with errors $T_{\epsilon_T} : \Omega \times A \to \Delta(\Omega)$ is defined by taking $T_{\epsilon_T}(\omega, a)(\omega')$ 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 $Q_{\epsilon_T,\epsilon_A} \in \mathcal{S}(\Omega, \Omega)$, the set of stochastic matrices, is the composition of the two transitions:

$$Q_{\epsilon_T,\epsilon_A}(\omega'; \omega) \equiv \sum_{a \in A} A_{\epsilon_A}(\omega)(\omega, a) T_{\epsilon_T}(\omega, a)(\omega') \tag{A-27}$$

We denote by $u_{\epsilon_A}(a)$ 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 ϵ_A. To illustrate, if the intended action profile is (C^1, D^2) then $u_{\epsilon_A}(C^1, D^2) = (1 - \epsilon_A)^2 s + \epsilon_A(1 - \epsilon_A)(d + c) + \epsilon_A^2 t$. We also let $V_{\epsilon_T,\epsilon_A}(\omega)$ the value function at the state ω

$$V_{\epsilon_T,\epsilon_A} = (1 - \delta) u_{\epsilon_A} + \delta Q_{\epsilon_T,\epsilon_A} V_{\epsilon_T,\epsilon_A} \tag{A-28}$$

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 as in the case of errors in transition only:

Lemma A.14. *The function $(\epsilon_T, \epsilon_A, \delta) \longrightarrow V(\cdot; \epsilon_T, \epsilon_A, \delta)$ 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 A.15. *The value function equation can be decomposed into nine independent equations, one for each of the invariant sets of the set Ω.*

1. $V_{\epsilon_T,\epsilon_A}(A, A) = u_{\epsilon_A}(D, D)$
2. $V_{\epsilon_T,\epsilon_A}(G, A) = (1 - \delta(1 - \epsilon_T)) u_{\epsilon_A} u(C, D) + \delta(1 - \epsilon) u_{\epsilon_A}(D, D)$
3. $V_{\epsilon_T,\epsilon_A}(A, G) = (1 - \delta(1 - \epsilon)) u_{\epsilon_A} u(D, C) + \delta(1 - \epsilon) u_{\epsilon_A}(D, D)$

Note that:

$$\begin{aligned} u_{\epsilon_A}(D, D) &= (1 - \epsilon_A)^2 d + \epsilon_A(1 - \epsilon_A)(s + t) + \epsilon_A^2 c \\ u_{\epsilon_A}(C, D) &= (1 - \epsilon_A)^2 s + \epsilon_A(1 - \epsilon_A)(d + c) + \epsilon_A^2 t \\ u_{\epsilon_A}(D, C) &= (1 - \epsilon_A)^2 t + \epsilon_A(1 - \epsilon_A)(d + c) + \epsilon_A^2 s \end{aligned}$$

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

Lemma A.16. *For all $\epsilon_T > 0$ and $\epsilon_A < 1/2$, (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:

$$V_{\epsilon_T,\epsilon_A}(A, A) - V_{\epsilon_T,\epsilon_A}(G, A) = (1 - 2\epsilon_A)(1 - \epsilon_A)(d + t - s - c).$$

□

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

A.3.1 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. Figure A.3 reports the sizes of the basin as function of the two error rates. These are the three dimensional versions of the Figure A.2.

One can consider also the exact analogue of Figure A.2, but for the error in action choice, by setting the transition probability to zero. This makes clear that the basin of attraction 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.

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.

A.4 Rational Learning Model

In this final section we show how the best response model can be interpreted as a model of Bayesian learning. This model will be used in section A.5 to analyze our data.

Each player has a belief on the distribution of strategies in the population. The belief has the same functional form for all players, a Dirichlet distribution of the three dimensional simplex, $\Delta(\{A, G, T\})$. A player with concentration parameter $\alpha \in \mathbb{N}^3$ has a density on beliefs given by:

$$D(\alpha^i) = \frac{1}{B(\alpha^i)} \mu(A)^{\alpha^i_A - 1} \mu(G)^{\alpha^i_G - 1} \mu(T)^{\alpha^i_T - 1} \quad \text{(A-29)}$$

where B is the multivariate beta function. The assignment of the belief described by an α 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.

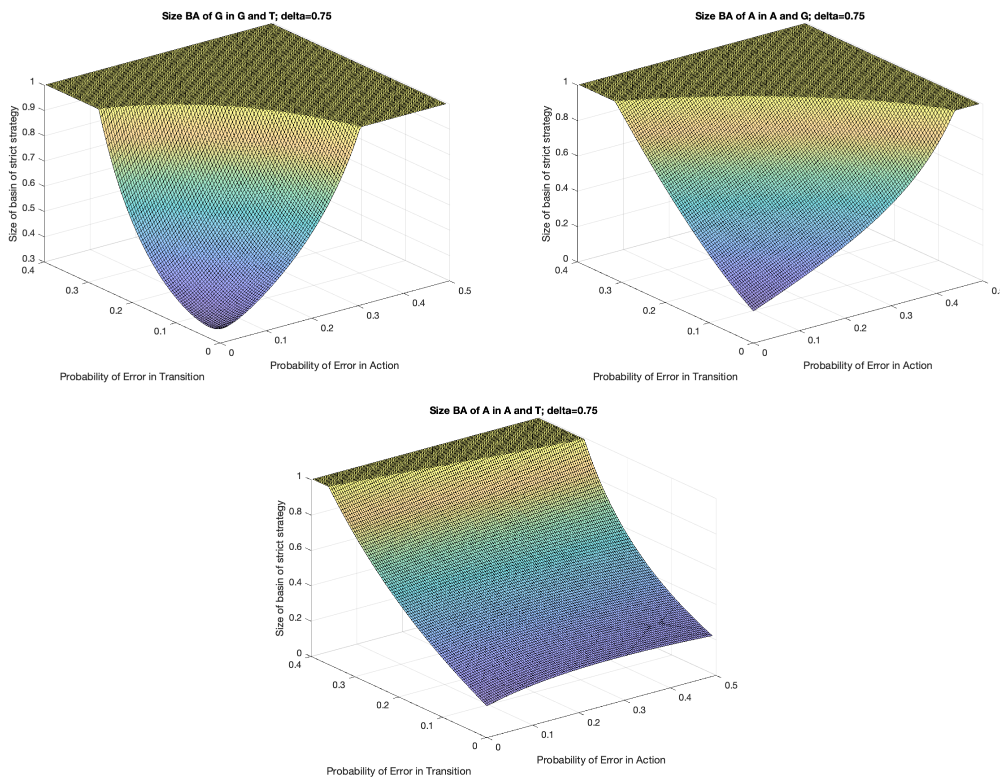
At time t, a fraction λdt of the population is randomly selected to adjust the strategy. This sub-population is representative of the total population, so the distribution on $\mathbb{N}^3$ in the sub-population is the same as in the total population. For convenience, we will consider in the following 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:

$$\pi \in \Delta(\mathbb{Z}^3) \quad \text{(A-30)}$$

We now consider players in the selected sub-population who are revising the strategy. Given the distribution on the strategy of the other selected players, each player computes and chooses a mixed strategy in the set of his best responses, given his belief $D(\alpha)$:

$$BR(\alpha) = argmax_{\sigma \in \Delta(M)} E_{D(\alpha)} U(\sigma, \cdot). \quad \text{(A-31)}$$

Figure A.3: **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 (G, T); A in (A, G); A in (A, T); respectively).



By the property of the Dirichelet distribution, the mean frequency of $m \in M$:

$$E_{D(\alpha)}\mu(m) = \frac{\alpha_m}{\sum_{r \in M} \alpha_m} \equiv \mathbf{E}(m; \alpha) \tag{A-32}$$

The best response set of a player with belief indexed by α is:

$$BR(\alpha) = argmax_{\sigma \in \Delta(M)} \sum_{r \in M} \mathbf{E}(m; \alpha) U(\sigma, m). \tag{A-33}$$

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.[36]

[36]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

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

Proposition A.17. *The fraction of each α belief in the model described in Section (A.4) follows the difference equation:*

$$\pi(k+1,\alpha) = S\pi(k,\cdot), k = 0, 1, \ldots. \tag{A-34}$$

where S defined in (A-36).

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

$$\alpha'_m = \alpha_m + \delta_m(b^i) \tag{A-35}$$

where δ_s is the indicator function.[37] Define:

$$(S\pi)(t,\alpha) \equiv \sum_{r\in M} \pi(t,\alpha - 1_m)\phi(m,\pi(t,\cdot)) \tag{A-36}$$

Recall our discussion before (A-30), so the definition (A-36) is meaningful even when $\alpha_m = 0$. Equation (A-34) follows. □

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 4.5, with the distribution on beliefs π replacing the distribution on strategies μ:

Proposition A.18. *The time derivative of the fraction of each α belief in the model described in section A.4 follows the equation:*

$$\frac{d\pi(t,\alpha)}{dt} = \lambda\left(\sum_{r\in M} \pi(t,\alpha - 1_m)\phi(s,\pi(t,\cdot)) - \pi(t,\alpha)\right) \tag{A-37}$$

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-35). 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 $\pi(t,\cdot)$ unchanged; and the sub-population of selected players, a fraction λdt, with frequency $S\pi(t,\cdot)$. Thus, the frequency next period is

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

and therefore (A-37) 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, $(\epsilon_T, \epsilon_A, \delta, U)$. 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.

sub-population of such players we get the expected value of the strategy choice.

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

A.5 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, $\alpha^i_{s,t}$.

A.5.1 Method

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, $t > 1$. We follow Dal Bó and Fréchette (2011), reporting here their setup for convenience, and assume that they update their beliefs as follows:

$$\alpha^i_{k,t+1} = \theta^i \alpha^i_{k,t} + 1(a^j_k), \tag{A-38}$$

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

$$U^i_{a,t} = \frac{\alpha^i_{A,t}}{\alpha^i_{A,t} + \alpha^i_{SC,t}} u_a(a^j_A) + \frac{\alpha^i_{SC,t}}{\alpha^i_{A,t} + \alpha^i_{SC,t}} u_a(a^j_{SC}) + \lambda^i \epsilon^s_{a,t} \tag{A-39}$$

where $u_a(a^j_k)$ is the payoff from taking action a when j takes the action k.

A.5.2 Estimation Results

The estimation of the model above generates choices of the first period of each supergame that in average fit well our data as shown in Figure A.4. We now analyze the two parameters of interest: θ^i, measuring the inverse of the speed by which subjects update their beliefs, and λ^i, measuring the inverse of the capacity of best responding given the beliefs.[38]

In Table A.5 we show the correlation between IQ and the parameters of interest. IQ significantly negatively correlated with θ^i, implying that higher IQ subjects update their beliefs faster. Higher intelligence does not significantly affect the capacity of best responding, modelled by the parameter λ^i. In the top panels of Figure A.5 we can compare the cumulative distribution of the θ^i in the different treatments. θ^i is smaller for high-IQ than for low-IQ, confirming that low-IQ update they beliefs slower than high-IQ (top left panel). The differences in the integrated treatment are drastically reduced (top right panel). Considering panel A of Table A.4 we note that the differences between high-IQ and low-IQ in the separated treatment is statistically significant, while the same difference in the integrated treatment is only weakly significant at the best. The bottom left panel

[38]Details on estimation procedures are in the online appendix of Dal Bó and Fréchette (2011) at page 6-8. We follow that procedure closely.

of Figure A.5 shows that low-IQ improve their speed (i.e. θ^i is lower) when integrated with the high-IQ, while there is no much difference among high-IQ subjects in the different treatments. Panel B of A.4 confirm that the differences among the low-IQ in the integrated and in the separated treatments are statistically significant. We can summarise this discussion as follows:

1. *higher intelligence subjects update their beliefs at a faster rate. Higher intelligence does not significantly affect the capacity of best responding;*

2. *less intelligent subjects learn to update their beliefs faster when they are mixed with more intelligent;*

3. *the way the subjects best respond to their beliefs does not depend on their intelligence.*

A possible explanation of why lower IQ subjects update their first period beliefs faster when mixed with the higher IQ is that, in the latter environment, they receive a clearer signal from the other players playing more consistent strategies.

Figure A.4: **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.

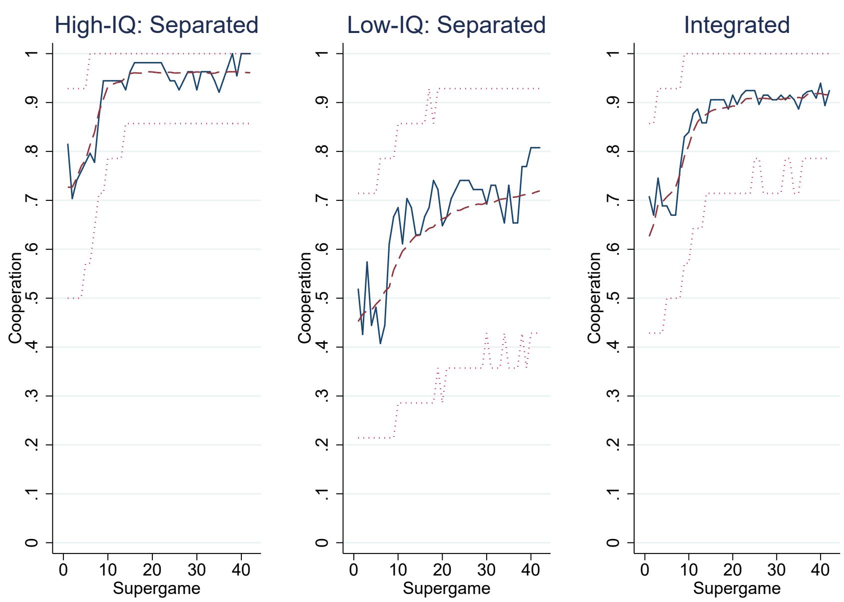


Figure A.5: **Distribution of the beliefs' updating speed within the different groups and treatments**. Distribution of the parameter θ^i as defined in equation A-38, where 1 correspond to slowest speed (fictitious play) and 0 to the fasted speed (Cournot dynamics)

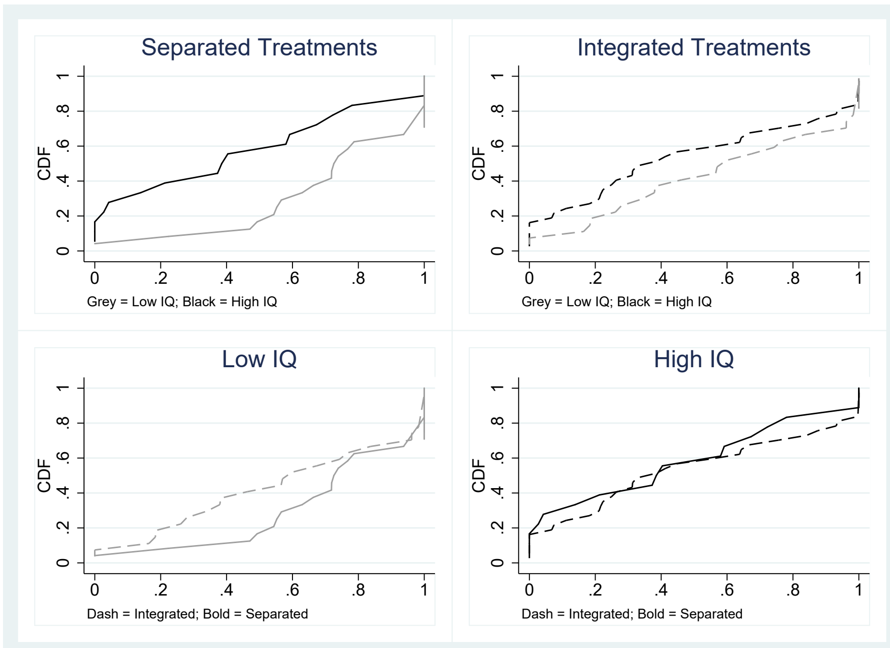


Table A.3: IQ and Simulated Parameters

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.4: **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-38, where 1 correspond to slowest speed (fictitious play) and 0 to the fastest speed (Cournot dynamics)
* $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

Panel A: Tests between IQ groups

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

Panel B: Tests between treatments

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

Table A.5: **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)	