Learning and Sophistication in Coordination Games

Kyle Hyndman · **Antoine Terracol** · **Jonathan Vaksmann**

June 29, 2009

Abstract This paper studies the role of strategic teaching in coordination games and whether changing the incentives of players to teach leads to more efficient coordination. We ran experiments where subjects played one of four coordination games in constant pairings, where the incentives to teach were varied along two dimensions – the short run cost of teaching and the long run benefit to teaching. We show which aspects of the game lead subjects to adopt long run teaching strategies, and show that subjects try to manipulate their opponent's actions to pull them out of a situation of coordination failure. We also show that extending a model of decision making by introducing a forward-looking component helps to track teachers' behaviour more accurately, and describes the way players behave in a more unified way across both teachers and learners.

JEL codes: C91, D43, D83

Keywords Game theory · Learning · Teaching · Coordination · Experiment

We would like to thank the Co-Editor, Jordi Brandts, and two anonymous referees whose suggestions have led to a substantially improved version. We are very indebted to participants at the ESA conference (Lyon), the Congress of the French Economic Science Association and seminar participants at Zaragoza Logistics Center for valuable comments. We also thank Maxim Frolov for his technical assistance. Financial support from SMU (URC Grant), the CNRS and Louis Lévy-Garboua (PPF project) is gratefully acknowledged. All remaining errors are our own.

Kyle Hyndman
Department of Economics, Southern Methodist University, 3300 Dyer Street, 301R, Dallas, TX 75275, hyndman@smu.edu

Antoine Terracol
EQUIPPE, Universités de Lille, and Centre d'Économie de la Sorbonne, Université Paris 1 - Panthéon Sorbonne, CNRS, terracol@univ-paris1.fr

Jonathan Vaksmann
Centre d'Économie de la Sorbonne, Université Paris 1 - Panthéon Sorbonne, CNRS, jonathan.vaksmann@univ-paris1.fr

1 Introduction

According to the original concepts of game theory, players are fully rational, which implies that in repeated games they completely anticipate the path of future play and take their actions accordingly. It is now commonly accepted that in many situations players' reasoning is more limited as their rationality might be bounded by their cognitive abilities. Several approaches introduce bounded rationality in the way players behave in repeated games. A large part of this literature — both theoretical and experimental — considers purely adaptive players.[1] In such learning models, players are modeled as myopic, taking actions based entirely on their past experience and under the assumption that their opponents' behaviour follows an exogenous process. Consequently, such myopic and adaptive players do not take into account the impact of their own actions on their opponents' future behaviour. In other words, according to these approaches, strategic interactions do not matter for players.

Even if full rationality might not seem reasonable in many cases, the assumption that strategic considerations do not play any role in repeated games also seems extreme in many situations. For example, Ellison (1997) studies a situation in which a single rational player is part of a large population of myopic players, with his main concern being when this lone rational player can move the population to a new equilibrium by acting non-myopically. He shows that the rational player can only move the population to a risk dominant equilibrium if he is sufficiently patient. Offerman et al (2001) has also noted that in too intricate games strategic reasoning is made very difficult and players consequently remain adaptive, but in other experimental environments, players have proven to be more sophisticated and use their actions, not only to optimize at a given time as myopic players would do but also to manipulate their opponents' behaviour in order to reach a preferable outcome in the future. In other words, players might attempt to teach their opponents.

We study teaching in coordination games. In particular, all of the games in our experiments have two Pareto rankable pure strategy Nash equilibria and one mixed strategy equilibrium. The equilibrium mixing probabilities are the same in all four games and are such that the inefficient equilibrium is risk dominant. Therefore, with myopic players we would expect frequent coordination failures.[2] The question we address in the present study is on the precise determinants of strategic teaching. In other words, we examine which payoff relevant elements are likely to trigger strategic behaviour from players and how it impacts the achieved outcome of a game. More precisely, teaching represents an investment according to which players might forego short-run payoffs by playing sub-optimal actions in order to manipulate their opponents and get more in the long-run by driving the outcome of a game towards the basin of attraction of a preferable equilibrium. Thus we design our games along two variables which parameterize both the short-run cost of teaching and the long-run gain of such a strategy.

[1] In microeconomics see, among others, Fudenberg and Levine (1998), Hopkins (2002), Erev and Roth (1998), Camerer and Ho (1999), Cheung and Friedman (1997), Crawford (1995), Samuelson (1998) and Weibull (1997). In macroeconomics see Marcet and Sargent (1989), Cho et al (2002) and Cho and Sargent (2008), among others.

[2] See, in particular, Ellison (1997), Kandori et al (1993) and Fudenberg and Levine (1998, Ch. 5).

To be sure, we are not the first to study strategic teaching experimentally. Ehrblatt et al (2008) and Terracol and Vaksmann (2009) identify the role of teaching on convergence to Nash equilibrium. They show that in fixed matching environments, teaching is relatively easily, which leads to higher convergence rates. On the other hand, when subjects are randomly matched or are given limited information about their opponent's payoffs, teaching is harder making convergence rare. However, neither study is particularly well-suited to inform upon which specific properties of the game are likely to cause teaches to emerge.[3] In the context of weak-link games, Brandts and Cooper (2006) and Brandts et al (2007) have shown that *leaders* often emerge who pull laggards out of coordination traps. In the former paper, the game is symmetric, so who the question of who *should* teach is difficult to determine, while in the latter paper, subjects within a group face different costs. Surprisingly, and unlike our results, it is not the subjects who have the lowest cost (and therefore the largest incentives to teach) that are the most likely leaders. Instead, leadership is driven by subjects with the most common cost type. Finally, in a different context, Cason et al (2008) demonstrates teaching in an indefinitely repeated game. Different from us, rather than teaching a stage game Nash equilibrium, players in their game teach an “alternation” strategy. Like us, they show that teaching changes as the degree on conflict in the game changes.

As previously mentioned, we vary the games according to the incentives of the row player to teach along two dimensions. In order to study sophistication, in addition to action choices, we also elicit beliefs of our subjects using a quadratic scoring rule. Unlike usual proxies for beliefs, we show that a subject's own action influences his stated beliefs about his opponent. This suggests that subjects believe their opponent is likely to best respond to their previously played actions, potentially making teaching possible. We term this phenomenon a *sophistication bias* in proxies for beliefs. We also show that the sophistication bias is greatest when the teaching incentives are the largest.

The above result about beliefs shows that teaching could be successful, but it doesn't show that subjects actually teach. Our next result is to demonstrate that subjects do, in fact, teach their opponent to play the Parteo efficient equilibrium when teaching incentives are strong. In contrast, when teaching incentives are weak, behaviour more closely mimics an adaptive learning rule. Beyond this, subjects' attempts at teaching are rewarded when the incentives are high — players more frequently coordinate on the efficient equilibrium and earn higher long run payoffs.

The paper is organized as follows. Section 2 introduces our games and experimental procedures. Section 3 shows that subjects converge more often to the efficient equilibrium when teaching incentives are high. Section 4 demonstrates sophistication in players' belief-formation process, while Section 5 analyses the determinants of sophistication in players' choice behaviour.

In Section 6 we consider a highly stylized empirical model which allows subjects to incorporate long run payoffs into their choice problem. Like Camerer et al (2002),

[3] In Ehrblatt et al (2008), there is a unique, Pareto efficient Nash equilibrium, making behaviour in the absence of teaching difficult to predict. On the other hand, the game used by Terracol and Vaksmann (2009) had three pure-strategy Nash equilibria with Pareto incomparable payoffs, making more than one teaching action quite likely.

our model presumes that sophisticated players believe that they can influence the actions of their opponent by repeatedly taking the same action. However, we differ in one important aspect. In their model, if a sophisticated player knows with certainty the learning parameters of his myopic opponent, then if teaching starts it can be expected to continue forever. We view teaching as a higher order learning process in which potential teachers learn about how quickly followers learn. In such a model, teachers could conceivably stop teaching after a number of periods if their opponent has not yet "caught on".[4] Our results show that when teaching incentives are high, subjects do incorporate long run payoffs and that our sophisticated model substantially outperforms the model with only myopic decision makers.

Finally, Section 7 concludes the paper.

2 Experimental Design

2.1 Games & Incentives

In order to examine the emergence of teaching, we conducted a number of experimental sessions. In particular, inexperienced subjects were brought into the experimental laboratory at the University of Paris 1 Panthéon-Sorbonne[5] and were asked to play one of the games in Figure 1 for a total of 20 periods. In order to give teaching the best shot at emerging, subjects were put in fixed pairs, and this information was clearly stated in the instructions. Before each experimental session began, subjects were randomly assigned the role of either a row or a column player and were told that they would remain in that role for the entire duration of the experiment. Payoffs were denominated in experimental currency units and were converted into Euros at the conclusion of the experiment, which generally lasted one hour or less. Subjects earned, on average, €14.1 for their participation.[6] A translation from the original French instructions given to the subjects can be found in an appendix at the end of the paper. In addition, subjects received an oral summary of the experimental conditions detailed in the instructions and questions were answered before the experiment began.

Figure 1 shows the games used in the experiment. Table headings are explained below in the text. Notice that all of our games have two pure-strategy Nash equilibria and one mixed-strategy Nash equilibrium. In these games, the pure strategy equilibria are Pareto rankable, both players strictly prefer the equilibrium (X,X) to the equilibrium (Y,Y). The mixed strategy equilibrium was $\{(0.8,0.2);(0.8,0.2)\}$. These kind of games are typically characterized by coordination failure and we will emphasize the determinants of strategic behaviour which might overcome this issue. Note that one desirable feature of our design is that, since all four games have the same mixed strategy equilibrium, their basins of attraction under adaptive learning must be largely

[4] That teachers might give up has been documented by Ehrblatt et al (2008). Such behaviour is also present in our data.

[5] For conducting the experiment we used the experimental software 'Regate' (Zeiliger, 2000).

[6] Throughout the paper, all payoffs are denominated in experimental currency units. The conversion factor was ECU100 = €2.1.

unchanged. Therefore, any differences across games in terms of coordination cannot be attributed entirely to adaptive behaviour.

Fig. 1 Payoff Matrices Used In The Experiments

$\text{TP}_h/\text{TC}_\ell$

	X	Y
X	40,45	8,37
Y	39,0	12,32

TP_h/TC_h

	X	Y
X	40,45	0,37
Y	37,0	12,32

$\text{TP}_\ell/\text{TC}_\ell$

	X	Y
X	20,45	8,37
Y	19,0	12,32

$\text{TP}_\ell/\text{TC}_h$

	X	Y
X	20,45	0,37
Y	17,0	12,32

It is natural to expect a teacher to try to teach his way to the Pareto optimal Nash equilibrium, (X,X). Likewise, in the absence of teaching, it is natural to expect frequent play of the risk-dominant equilibrium, (Y,Y).[7] In order to systematically study teaching, we originally conjectured that two parameters that would affect teaching can be used to describe coordination games. Each of the games employed, therefore, varies one of these parameters which we now explain in some detail. More precisely, teaching is best thought of as an investment: the successful teacher will incur short-run costs in order to obtain a long-run gain. Therefore, the games were designed to vary either the short-run cost or the long-run gain. Note, however, that in order to study teaching, we also need teachers to be paired with subjects who are capable of being taught (*e.g.,* an adaptive learner). In order to do this, in all of our games we kept the payoffs to the column player fixed and, moreover, the incentives that the column player had to engage in long-run behaviour were always lower than those of the row player. As for the row players, their incentives to teach were varied from low to high in each of two dimensions which we now discuss.

First, consider the short-run costs associated with teaching: by playing X when he believes that his opponent will play X with probability p, player i, $i = r,c$, incurs a *teaching cost* (TC), given by

$$E_i^Y(p) - E_i^X(p) = \theta_i \cdot (p^* - p)$$

where $E_i^a(p)$ is player i's expected payoff from taking action a, $a = X,Y$, given a belief of p, p^* is the equilibrium mixing probability (*i.e.,* $p^* = 0.8$) and, denoting $\pi_i(a,a')$ player i's payoff when he plays a and his opponent plays a', we have:

$$\theta_i = \pi_i(X,X) - \pi_i(X,Y) + \pi_i(Y,Y) - \pi_i(Y,X)$$

Thus, θ indexes a player's teaching cost and can be called the teaching cost parameter. The teaching cost is thus a penalty attached to playing a teaching action despite the

[7] Refer to the citations mentioned in Footnote 2 for theoretical justification. The experiments of Battalio et al (2001) documented frequent play of the inefficient equilibrium. In weak link games, van Huyck et al (1990), Knez and Camerer (1994, 2000), Brandts and Cooper (2006), Brandts et al (2007) and Chaudhuri et al (2009), among others, showed frequent play of inefficient equilibria.

fact that it is not a best response to static beliefs. We expect teaching to be negatively correlated with the teaching cost parameter since higher short-run costs of teaching should focus attention on static profit maximisation.

Next consider the long-run return on investment. By successfully teaching, the game converges to (X,X). So teaching implies playing X and facilitates the emergence of the good equilibrium (X,X), even though playing Y would be a best response but would facilitate the emergence of the bad equilibrium (Y,Y). Therefore, we will measure the *teaching premium* (TP) as the percentage increase in payoffs going from the bad to the good equilibrium. That is, the teaching premium parameter for player i is:

$$\psi_i = \frac{\pi_i(X,X) - \pi_i(Y,Y)}{\pi_i(Y,Y)}$$

and we expect the teaching premium parameter to be positively correlated with teaching since a higher teaching premium implies a higher long-run reward from teaching.

We ran four different games varying the size of the teaching cost and the teaching premium parameters, High or Low, for the row player. In order to encourage column players to remain passive, we kept their teaching incentives constant and substantially weaker than for the row player. We expect row players, particularly when the incentives are strong, to take the role of a teacher. We summarize the experimental games in Table 1.

Table 1 Summary of Experimental Games

Game	ψ_r	θ_r	ψ_c	θ_c	#Subjects
$\text{TP}_h/\text{TC}_\ell$	2.33	5	0.41	40	34
TP_h/TC_h	2.33	15	0.41	40	32
$\text{TP}_\ell/\text{TC}_\ell$	0.67	5	0.41	40	38
$\text{TP}_\ell/\text{TC}_h$	0.67	15	0.41	40	30

Remark 1 Observe that what we call the teaching cost is simply what Battalio et al (2001) have called the "optimization premium". These authors focus on adaptive (myopic) behaviour in 2×2 coordination games with Pareto rankable equilibria and find that the lower is the optimization premium, or in our view, the lower is the teaching cost, the more efficient is coordination. Although Battalio et al (2001) consider a random matching environment which mitigates the incentives to teach, their games confound the optimization and teaching premia; in particular, as they lower the optimization premium from game to game, they simultaneously increase the teaching premium, potentially making attribution of cause problematic. In our experiments, we try to isolate both of these effects separately. Moreover, our fixed-pairs matching protocol should give our subjects pause to consider the long-run benefits of teaching — in our terminology, the teaching premium — and so we also look for sophistication in subjects' behaviour by studying a particular forward-looking model of decision making (see Section 6).

2.2 Belief Elicitation

In this study, among other things, we aim to perform a detailed examination of players' propensity to play sub-optimal actions during a possible teaching phase. In order to do this, we must elicit players' beliefs to precisely determine their best response at each time. In each round, before choosing their action, subjects also reported their beliefs about the likely action of their match in that round. Beliefs were rewarded for accuracy according to a quadratic scoring rule, which should induce truth-telling if subjects are risk neutral.[8] The exact parameterisation of the QSR used can be found in the experimental instructions. As usual in this kind of design, we tried to keep the reward for reporting beliefs small in comparison with the payoffs associated to the game so that players could not use their belief payoff as a "hedge" against potentially low stage-game payoffs.

At the end of each round, subjects were informed about the action of their opponent, their game payoff, their prediction payoff and the game payoff of their opponent for the current round. When deciding in a given round, subjects could always see on their screen the entire history of actions and stage game payoffs as well as their predictions in earlier rounds, although they could not see their prediction payoffs from earlier rounds.

3 Summary: Convergence & Coordination

We begin our analysis of the experimental results with a brief look at the outcomes of the games that our subjects played. Specifically we look at how well coordinated subjects' actions were, including whether or not they converged and, if so, which equilibrium they converged to.

Table 2 tabulates the number of groups that converged to each of the pure equilibria as well as those who did not converge. We will say that the game converged to a pure strategy Nash equilbrium if both players chose their part of the Nash equilibrium for *at least* 3 consecutive periods before the end of the game (*e.g.,* periods 18, 19 and 20).[9] For each game, we also conduct a proportions test to see whether there is a difference in convergence to the efficient equilibrium vs. the inefficient equilibrium. For the game $\text{TP}_h/\text{TC}_\ell$, three times as many groups converge to the efficient equilibrium as do to the inefficient equilibrium. Moreover, this difference is significant at the 5% level. A similar result holds for the game $\text{TP}_\ell/\text{TC}_\ell$, though the difference is only significant at the 10% level. Notice that only in the game least conducive to teaching

[8] Several studies (e.g. Offerman and Sonnemans (2001), Nyarko and Schotter (2002)) indicate that subjects report their true beliefs when incentivized by the Quadratic Scoring Rule. Rutström and Wilcox (2004), however, finds that an intrusive scoring rule for belief elicitation affects people's behaviour. More recently, Costa-Gomes and Weizsäcker (2008) report that subjects choose actions in one-shot games assuming a very low level of rationality of their opponent, but that their beliefs are generally more sophisticated. In the language of level-*k* theory, subjects play according to L1 but report L2 beliefs.

[9] We will also allow games with a final period deviation by one of the players to be considered convergent, but in such cases we require that both players chose the Nash action for 5 or more periods before the final period deviation.

(*i.e.,* $\text{TP}_\ell/\text{TC}_h$) do more groups converge to the inefficient equilibrium, although the difference is not statistically significant.

Table 2 Tabulating convergence in our games

	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$
(X,X)	9^{**}	6	8^{*}	4
(Y,Y)	3	4	3	6
N.C.	5	6	8	5

** significant at 5%; * significant at 10%

As teaching is *per se* a dynamic strategy, Table 3 shows the average frequency subjects coordinated on (X,X) and (Y,Y), respectively, in the first 10 and in the last 10 periods. As can be seen, for the game $\text{TP}_h/\text{TC}_\ell$ our subjects played the Pareto efficient equilibrium more frequently in the latter half of the game, while for the games $\text{TP}_\ell/\text{TC}_h$ and TP_h/TC_h they played the Pareto *inefficient* equilibrium significantly more frequently in the latter half of the game. Interestingly, we see that despite the weak incentives to teach, subjects in the $\text{TP}_\ell/\text{TC}_\ell$ game managed to play the Pareto efficient equilibrium approximately 45% of the time — even in periods 1 – 10, they played it 44% of the time.

Table 3 An Examination of Successful Coordination

	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$
$(X,X)_{1-10}$	0.300	0.375	0.437	0.327
$(X,X)_{11-20}$	0.471	0.425	0.463	0.327
paired t-test	1.84^{*}	1.14	0.43	0.00
$(Y,Y)_{1-10}$	0.253	0.231	0.274	0.267
$(Y,Y)_{11-20}$	0.235	0.319	0.284	0.487
paired t-test	0.38	1.78^{*}	0.14	2.55^{***}

* 10% level of significance; ** 5% level of significance; *** 1% level of significance. For each statistic the number of observations is the number of pairs in each game, i.e. respectively 17, 16, 19 and 15 for games $\text{TP}_h/\text{TC}_\ell$, TP_h/TC_h, $\text{TP}_\ell/\text{TC}_\ell$ and $\text{TP}_\ell/\text{TC}_h$.

Finally, if we look at the frequency of efficient coordination (*i.e.,* the fraction of times players coordinated on (X,X), conditional on playing (X,X) or (Y,Y)), we see that this fraction is significantly higher for the game $\text{TP}_h/\text{TC}_\ell$ than for the game $\text{TP}_\ell/\text{TC}_h$ in the second half of the game ($z = 1.711$, $p = 0.097$). In no other cases was there a significant difference. That is, at least when the discrepancy between teaching incentives is particularly large, players are significantly more able to overcome the coordination failure and more often reach the Pareto efficient equilibrium.

Thus, the above-mentioned results on coordination across games are broadly consistent with our conjectures about players' teaching incentives. Moreover, remember that, as we said in Section 2, these differences in the equilibria achieved are hard to rationalize under the perspective of purely adaptive learning since, because of the symmetry in the (pure and mixed) equilibrium structures of our games, we do not expect the forces that drive players' behaviour under these approaches to make different outcomes emerge. We believe that these descriptive results suggest a more sophisticated view of players' behaviour, that we will precisely examine in the rest of the paper.

4 Do Subjects View Their Opponents as Being Capable of Learning?

The above results show that when teaching incentives are high, players are more likely to converge to the Pareto efficient equilibrium, while when teaching incentives are low, they are more likely to converge to the Pareto inefficient equilibrium. We now seek to show that these differences can be explained by differences across treatments in the degree of strategic teaching. Note that successful teaching requires a partner who is capable of learning and a teacher who believes that his partner is, in fact, capable of learning. Indeed, if the teacher (he) believes that his opponent (she) updates his actions largely due to the observed history of play, and that she updates sufficiently rapidly, then he might be willing to make the required short term investment. In this section, we examine whether players beliefs are influenced by their own actions; that is, whether they see their opponents as learners.

Our strategy is to determine whether subjects *believe* that they can influence their opponent's actions through their own choices. This implies, in the spirit of Terracol and Vaksmann (2009), an investigation of players' belief formation process to check whether they take into account the influence brought by their own past actions when forming beliefs about their opponents' behaviour at a given time. That is, we want to see whether subjects view their opponent as an adaptive learner who is capable of being taught something.

Our aim in this section is not to explicitly model the way a player's action impacts his opponent's behaviour. Instead, we will show that a player's beliefs depend also upon their own actions (in addition to the actions of one's opponent), and so, are more sophisticated than is often assumed, which validates a necessary pre-condition for teaching. More precisely, before players *teach* strategically, they first have to *think* strategically and perceive that their own past actions are likely to influence their opponent's current and future behaviour. A naive way to check this would be to directly examine whether players' beliefs vary according to their own previously chosen action. Beliefs, however, may also depend on the history of the opponents' past actions, as postulated by traditional proxies used to describe players' belief-formation process. If it is the case one must filter out the impact of these past actions to avoid spurious correlations between $a_i(t-1)$, player i's own action in the previous round, and his current beliefs.[10]

[10] If players partly base their beliefs on the past history of their opponents' play, then this component will be correlated to both their beliefs in the previous and current rounds, and also to their action in the previous round (because $a_i(t-1)$ obviously depends on player i's beliefs in $t-1$), and one would find a positive correlation between current beliefs and previous action *even if* there is no causal effect of the previous action on the current belief. Note that we do not assume that players necessarily base their beliefs on the history of their opponents' play, but rather allow for the possibility of such a belief formation process.

Adopt the terminology of Nyarko and Schotter (2002) and refer to beliefs based only on the history of the opponents' actions as "empirical" beliefs. Denote empirical beliefs by $\tilde{B}_i^a(t)$ to distinguish from the stated beliefs, $B_i^a(t)$, reported by players. Next define $D_i^a(t) = B_i^a(t) - \tilde{B}_i^a(t)$ to be the difference between stated and empirical beliefs. Again, taking this difference allows us to avoid the above-mentioned spurious correlations. Observe that since $\tilde{B}_i^a(t)$ is conditional on the history of the opponents' past actions, but not on the action chosen by player i in period $t-1$ (*i.e.,* $a_i(t-1)$), then if $D_i^a(t)$ depends on $a_i(t-1)$, so too must stated beliefs. In this case, we may conclude that players think that their opponents modify their behaviour according to the history of the game, and take this into account in their own beliefs; in other words subjects realize that their opponents can learn, which is necessary for teaching to even be possible.

We chose to model empirical beliefs, $\tilde{B}_i^a(t)$, with the γ-weighted beliefs model of Cheung and Friedman (1997), where the belief held by player i about the probability that player j will play action a in round $t+1$ is given by:

$$\tilde{B}_i^a(t+1) = \frac{\mathbb{1}_{\left(a_j(t)=a\right)} + \sum_{u=1}^{t-1} \gamma^u \mathbb{1}_{\left(a_j(t-u)=a\right)}}{1 + \sum_{u=1}^{t-1} \gamma^u} \tag{1}$$

where $\mathbb{1}_{\left(a_j(t)=a\right)}$ equals one if player j has played action a in round t, and zero otherwise. Actions played in a given round are discounted with time at rate $\gamma \in [0,1]$. When $\gamma = 0$, this model reduces to Cournot learning, where the belief held in period t about action a is one if the action has been play in round $t-1$, and zero otherwise, while when $\gamma = 1$, the model reduces to fictitious play, where the belief about a given action corresponds to the frequency with which this action has been played since round 1. The Cheung and Friedman model has been found to perform well empirically to explain people's behaviour in games. Again, according to these empirical beliefs, players form conjectures only on the basis of the history of their opponent's actions, but they do not realize that their own actions are likely to influence their opponent's behaviour. In other words, players regard their opponent's behaviour as generated by an exogenous process and, in doing so, they completely neglect strategic interactions. This assumption is at odds with the foundations of game theory. We now examine its validity in our games.

We estimate the model of equation (1) at the individual level using the method of minimum mean-squared error[11] along the lines of Nyarko and Schotter (2002). We are thus able to compute estimated empirical beliefs $\hat{\tilde{B}}_i^a(t)$ that can be interpreted as the largest part of the individual's true beliefs $B_i^a(t)$ that can be explained by the past history of the opponents' actions up to round $t-1$ under the Cheung-Friedman hypothesis. We then compute $\hat{D}_i^a(t)$, as the difference between stated beliefs and estimated empirical beliefs in round t. We then proceed to examine whether $\hat{D}_i^a(t)$ can be explained, in part, by the action taken by player i in the previous round — $a_i(t-1)$. Specifically, we estimate the following random-effects model:[12]

$$D_i^X(t) = \beta_0 + \beta_1 \tilde{B}_i^X(t) + \beta_2 \mathbb{1}_{(a_i(t-1)=X)} + v_i + \varepsilon_{i,t} \tag{2}$$

[11] That is, our estimator is the solution to $\min_{\gamma\in[0,1]} \sum_{a,t} \left(B_i^a(t) - \tilde{B}_i^a(t)\right)^2$.

[12] By complementarity, results are the same for action Y.

The variable $\tilde{B}_i^X(t)$ serves as a control variable as the size of the difference $D_i^X(t)$ will also depend on the value of player i's empirical beliefs since a high empirical belief leaves less room for a large positive $D_i^X(t)$ than low empirical beliefs. Now consider β_2. If $\beta_2 > 0$, then i believes that his opponent is more likely to best respond to the previous action than implied by γ-weighted beliefs and we say that subject i believes that his opponent is capable of learning. Hence this indicates that players (at least partly) base their actions on motivations beyond those suggested by classical adaptive proxies. For this reason, we will refer to it as a *sophistication bias*.

Our estimation results of Equation (2) for the variable of interest $\mathbb{1}_{(a_i(t-1)=X)}$ are collected in Table 4, for all players and separating by type of players (row or column) respectively. We always find a significantly positive parameter β_2. The results indicate that players are in fact more likely to think that their opponent will choose a best response to their previous action than implied by usual proxies for empirical beliefs.[13] In other words, subjects realize that their opponents can learn, which is necessary for teaching to even be possible. This shows that players take strategic interactions into account and form beliefs in a more sophisticated way than the adaptive way postulated by usual proxies. Therefore we have highlighted a *sophistication bias* in classical proxies used to describe players' belief-formation process.

Table 4 Random-Effects Panel Regression: The Sophistication Bias

	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$
All	0.149^{***} (0.032)	0.210^{***} (0.041)	0.137^{***} (0.042)	0.187^{***} (0.062)
Row players	0.138^{***} (0.046)	0.230^{***} (0.068)	0.167^{**} (0.065)	0.173^{*} (0.091)
Column players	0.163^{***} (0.045)	0.195^{***} (0.049)	0.098^{**} (0.048)	0.199^{**} (0.086)

* 10% level of significance; ** 5% level of significance; *** 1% level of significance
Robust standard errors in parentheses. The number of individuals is given in Table 1, each individual played 20 periods.

When separating by type of players, all coefficients β_2 remain positive and significant, indicating that our finding is robust in our games as both types of players take strategic interactions into account when forming their beliefs. In sum, these results emphasize the fact that subjects believe that their past actions influence their opponent's current decisions — i.e., there is a sophistication bias. This finding is particularly interesting for our purpose since it means that a crucial pre-condition for successful teaching is met: players realize that their opponent is capable of learning. Of course, the presence of a sophistication bias does *not* allow us to conclude that subjects actually teach their opponent. In the next section, we show that, in addition to having a more sophisticated belief-formation process, subjects' action choices are also considerably more sophisticated than traditionally postulated.

[13] β_1, the coefficient of the control variable is consistently always significantly negative as, again, a higher level of estimated beliefs leaves less room for large positive differences.

5 Over response & Teaching

The previous section confirmed that in all our games there is scope for teaching. The question we address now is whether subjects, particularly row players given their stronger incentives, take advantage of this scope and actually attempt to teach their way to a beneficial (Pareto optimal) outcome.

In the belief-learning literature, players are assumed to take a stochastic best response to their beliefs. Under this view, choices which are not a best response to beliefs are called errors. However, if agents are teaching, then they are necessarily taking a statistically sub-optimal action with the hope that a more beneficial outcome will emerge later. In order to capture this, in Table 5 we categorise the players' choices according to whether or not they were a best response to stated beliefs. Say that a player "over responds to X" whenever he chooses X despite the fact that Y is a best-response to his stated beliefs. Similarly, say that a player "over responds to Y" whenever he chooses Y despite the fact that X is a best-response to his stated beliefs. If our subjects are teaching in these games, we would expect the players to over respond to X much more frequently than they over respond to Y. Indeed, this is precisely what we see: as can be seen from Table 5, in no case were more than 6% of the decisions choices of Y when X was a best response, while X was played when Y was a best response for at least 17% of the decisions and often for over 20% of the decisions (this is seen by looking at the off-diagonals of the table). A bit more formally, if we take each subject-pair's 20 period history as an independent observation, we can carry out a two-sided t-test of equality of means. For row players in the game $\text{TP}_\ell/\text{TC}_h$ (*i.e.* the game which exhibits the weakest teaching incentives) the p-value of this test is only 0.021, while in all other cases, $p \ll 0.01$.[14]

It is hard to interpret such strong tendencies to choose X even when Y is a best response as errors since it would mean that our subjects are making quite costly errors with considerable frequency. Moreover, the comparative statics are consistent with our earlier hypotheses. When the teaching premium is low and the teaching cost is high, the difference in over response behaviour is much weaker, though the difference is still statistically significant. In contrast, when the teaching incentives are higher, Table 5 shows that our subjects choose X more frequently when Y is a best response than the converse.

As we said in Section 2, we purposely gave the column players very weak incentives to teach, while giving the row players stronger incentives to teach. Therefore, if our conjectures are correct, it should be the case that row players over respond to X more frequently than do column players. Indeed, when teaching incentives are highest, this is the case. In the game $\text{TP}_h/\text{TC}_\ell$, a two-sided t-test of the null hypothesis that row and column players over respond to X with the same frequency is rejected ($z = 2.25$, $p = 0.03$). When teaching incentives are weaker, we can never reject the same hypothesis (in all cases, $p > 0.2$).

[14] In order to remain conservative, when players were indifferent between their two actions and finally chose X, we regarded this action as a best response and not as an over response. It turns out that whether we consider this as a best response or an over response does not alter our results as these cases are relatively rare occurrences.

Table 5 Frequency of Choice Behaviour Categorised By Best Response

ROW PLAYERS

$\text{TP}_h/\text{TC}_\ell$

	$BR = X$	$BR = Y$
X	0.25	0.38
Y	0.02	0.36

TP_h/TC_h

	$BR = X$	$BR = Y$
X	0.31	0.26
Y	0.01	0.42

$\text{TP}_\ell/\text{TC}_\ell$

	$BR = X$	$BR = Y$
X	0.37	0.23
Y	0.04	0.36

$\text{TP}_\ell/\text{TC}_h$

	$BR = X$	$BR = Y$
X	0.29	0.17
Y	0.06	0.48

COLUMN PLAYERS

$\text{TP}_h/\text{TC}_\ell$

	$BR = X$	$BR = Y$
X	0.27	0.24
Y	0.04	0.45

TP_h/TC_h

	$BR = X$	$BR = Y$
X	0.37	0.19
Y	0.02	0.43

$\text{TP}_\ell/\text{TC}_\ell$

	$BR = X$	$BR = Y$
X	0.39	0.18
Y	0.03	0.40

$\text{TP}_\ell/\text{TC}_h$

	$BR = X$	$BR = Y$
X	0.29	0.20
Y	0.04	0.47

The numbers in each matrix should sum to 1, modulo rounding.

Finally, it should be the case that, at the very least, row players teach more in the game $\text{TP}_h/\text{TC}_\ell$, while there should be no significant difference in teaching behaviour across games for the column players. This turns out to be the case. In Table 6, which provides the t-statistics for every two-sample test that subjects in one game choose X when Y is a best response with the same frequency as in another game. A positive value in a given cell indicates that the subjects in the game along the row chose X when Y is a best response more frequently than subjects in the game along the column, and vice-versa for negative t-statistics. As the reader can see, row players in the game $\text{TP}_h/\text{TC}_\ell$ chose X when Y was a best response significantly more frequently than did subjects in all other games. This result reinforces our belief that these are not errors that we are picking up, but rather the deliberate attempt of row players to teach their way to the Pareto efficient equilibrium. We also see that row players in the TP_h/TC_h game chose X when Y was a best response significantly more frequently than did subjects in the $\text{TP}_\ell/\text{TC}_h$ game. In no other case is there a significant difference in behaviour for row players. This suggests to us that the teaching premium is the more salient variable when it comes to inducing teaching. Briefly, looking at the table for column players, we see that there are no significant differences in behaviour between games.

Next observe that since teaching involves incurring a short-run cost for a long-term gain, it should also be the case that teaching diminishes as the game progresses. Therefore, to the extent that our proxy (*i.e.,* choosing X when Y is a best response) captures teaching, we should see that such behaviour declines in later rounds. In fact, this is exactly what happens in all games, though there are two important points to make. First, in all treatments and for both row and column players the frequency of over-response declines as the game proceeds. This is likely due to some combination of learning and the fact that teaching becomes less and less beneficial as the

Table 6 Two-sample t-tests Across Games: Frequency of X Choices When Y is a Best Response

Row Players

	$\mathrm{TP}_h/\mathrm{TC}_\ell$	$\mathrm{TP}_h/\mathrm{TC}_h$	$\mathrm{TP}_\ell/\mathrm{TC}_\ell$	$\mathrm{TP}_\ell/\mathrm{TC}_h$
$\mathrm{TP}_h/\mathrm{TC}_\ell$	-	1.75^{*}	2.79^{***}	4.19^{***}
$\mathrm{TP}_h/\mathrm{TC}_h$	-	-	0.83	2.03^{**}
$\mathrm{TP}_\ell/\mathrm{TC}_\ell$	-	-	-	1.26
$\mathrm{TP}_\ell/\mathrm{TC}_h$	-	-	-	-

Column Players

	$\mathrm{TP}_h/\mathrm{TC}_\ell$	$\mathrm{TP}_h/\mathrm{TC}_h$	$\mathrm{TP}_\ell/\mathrm{TC}_\ell$	$\mathrm{TP}_\ell/\mathrm{TC}_h$
$\mathrm{TP}_h/\mathrm{TC}_\ell$	-	0.94	1.52	0.56
$\mathrm{TP}_h/\mathrm{TC}_h$	-	-	0.54	0.30
$\mathrm{TP}_\ell/\mathrm{TC}_\ell$	-	-	-	-0.79
$\mathrm{TP}_\ell/\mathrm{TC}_h$	-	-	-	-

* 10% level of significance; ** 5% level of significance; *** 1% level of significance
A positive (negative) t-statistic indicates that the subjects in the game along the **row** chose X when Y was a best response more (less) frequently than did subjects in the game along the **column**. There are respectively 17, 16, 19 and 15 row (column) players in games $\mathrm{TP}_h/\mathrm{TC}_\ell$, $\mathrm{TP}_h/\mathrm{TC}_h$, $\mathrm{TP}_\ell/\mathrm{TC}_\ell$ and $\mathrm{TP}_\ell/\mathrm{TC}_h$.

game proceeds. Second, row players in the game $\mathrm{TP}_h/\mathrm{TC}_\ell$, retain a higher proportion of over responses throughout the game and the proportion decreases more slowly. For row players in the game $\mathrm{TP}_h/\mathrm{TC}_\ell$, the proportion of over-response declines most rapidly. For column players, there are no noticeable differences across treatments.

Our results thus far provide support for our main hypothesis: subjects respond in predicted ways to changing incentives to teach. When teaching is not costly and/or the benefit to successful teaching is high, they teach. When teaching is costly and/or the benefit is low, they do not. This is seen in both row and column players: generally, row players engage in what looks like strategic teaching more frequently than do column players, which is natural since their incentives were always larger. This also suggests that column players took a more passive role and were more likely to be *followers*. Moreover, row players respond to the changing teaching incentives across games. The dynamics of over response are also consistent with teaching as it generally decreases with time and particularly more slowly when teaching incentives are the highest.

6 A Demonstration of Far-sighted Behaviour

The previous sections have provided strong evidence of sophisticated behaviour in our games. In this section, we attempt to track this pattern. To do so we write down and then estimate a descriptive model that aims to capture some prominent behavioral traits. As will be seen, the results of this exercise generate comparative statics which are consistent with our conjectures and our previous results.

6.1 Motivation

Like many others, we start from the premise that some players may engage in far-sighted behavior.[15] The theoretical literature on sophisticated decision makers demonstrates that their presence may lead to Stackelberg payoffs for the sophisticated player (Fudenberg and Levine, 1989), may lead to the risk dominant equilibrium (Ellison, 1997), may force cooperation (Jehiel, 2001) or may lead to efficient outcomes — not necessarily equilibria (Mengel, 2008), among other results.

Using experimental data, Camerer et al (2002) has empirically documented the existence of sophisticated agents who, through their own actions, try to manipulate the attractions of other players in order to maximise their long run payoff. One important assumption in this model is that the sophisticated players know the precise learning model used by the myopic players.[16] It is this assumption that we take issue with.

To see why, consider a sophisticated agent matched with a myopic player who learns according to the Cheung and Friedman (1997) model of γ-weighted beliefs. For ease of exposition, suppose that the myopic player perfectly best responds to his beliefs and has an initial weighting $(w_0^X, w_0^Y) = (0,1)$, so that he believes with probability 1 that the sophisticated player will choose action Y, making it a best response to choose action Y as well. Next suppose that the sophisticated player believes that the $\gamma = 0.7$ for the myopic player. Then, after one period of playing X, the myopic player's beliefs are approximately 0.588; after two periods they are approximately 0.776 and after three periods they are approximately 0.813. Therefore, given the sophisticated player's beliefs about the learning parameter of the myopic player, by the fourth period, the myopic player's beliefs *should be* such that X is a best response. How should the sophisticated player react if the myopic player chose Y in the fourth period? In our view, he must reevaluate his assumption that $\gamma = 0.7$ in favour of some new belief that γ is actually higher.[17] Given this new belief about γ, continued teaching may no longer be optimal. Indeed, Ehrblatt et al (2008) document that teachers will often stop teaching if their opponent is not responsive enough.

Given this motivation, we seek a model which captures the above intuition; that is, some players may contemplate playing a teaching strategy to manipulate the beliefs of their opponent. Throughout this process, teachers reevaluate the speed at which their opponent learns, and decide whether continued teaching is optimal. The model we present below is highly stylized, making a number of simplifying assumptions, but is flexible enough to meet our requirements.

[15] See, for example, Fudenberg and Levine (1998, Ch. 8), and Camerer et al (2002), among others. See also the level-*k*/cognitive hierarchy models of Stahl and Wilson (1995), Costa-Gomes et al (2001) and Camerer et al (2004).

[16] However, they may be mistaken about the proportion of sophisticated players in the population.

[17] To be sure, the sophisticated agent may also reevaluate the assumption that the myopic player perfectly best responds to his beliefs, or some combination of both. In general, updated beliefs about the learning parameters of the myopic player would be towards less rationality and greater sluggishness in beliefs.

6.2 The model

The key features of the model are:

1. Sophisticated players see their opponent as a γ-weighted belief learner à la Cheung and Friedman (1997) and contemplate playing a teaching strategy in order to manipulate their opponent's beliefs.
2. Sophisticated players maximize the discounted sum of expected payoffs.
3. Sophisticated players do not know the true value of γ of their opponent, but update it based on the observed actions of their opponent.

The assumption that sophisticated players view their opponent as a γ-learner means that we can restrict attention to two continuation strategies: repeatedly playing X or repeatedly playing Y.[18] Repeatedly playing X is done in an effort to push play to the Pareto efficient equilibrium. We call this the teaching strategy.

For $A \in \{X,Y\}$, let $\sigma_i^A(t) = (a_i(t),\ldots,a_i(T)) = (A,\ldots,A)$ denote a continuation strategy of the sophisticated player i at time t. Such a player seeks to maximize his expected intertemporal payoff. Thus, he chooses $\sigma^A(t)$ to maximize:

$$\begin{aligned} E_i(\sigma^A(t)) = {} & b_i^X(t)\cdot\pi_i(A,X) + (1-b_i^X(t))\cdot\pi_i(A,Y) \\ & + \textstyle\sum_{u=t+1}^{T} \delta^{u-t} \sum_{z=X,Y} b_i^z(u|\sigma^A(t))\cdot\pi_i(A,z), \end{aligned} \tag{3}$$

where $\pi_i(a,a')$ is player i's payoff when he plays a and his opponent plays a'. $b_i^z(t)$ is player i's current (i.e. time t) belief about action z, $z = X,Y$. $b_i^z(u|\sigma^a(t))$ is player i's "prospective belief" at time u, $u > t$, about action z conditional on the fact that he adopts the continuation strategy $\sigma^a(t)$ from t until the end of the game. δ is the usual discount factor, and observe that if $\delta = 0$, then player i is actually myopic.

δ is a parameter to be estimated. Therefore, it remains to specify how current and prospective beliefs are formed and how player i maps expected payoffs into actions. We turn to this now.

The Opponent's Beliefs. As indicated above, the sophisticated player i regards his opponent as a γ-learner but does not know the true value of γ. We assume that player i updates the value of the inertia parameter $\hat{\gamma}_j(t)$ he perceives on the basis of the information he gathered at time t.[19] The empirical rule we adopt to describe players' updating process of $\hat{\gamma}_j(t)$ is the following:

$$\hat{\gamma}_j(t) = \left\{ \begin{array}{cc} 0 & \text{if } t \leq 2 \\ \frac{\hat{\gamma}_j(t-1)+\sum_{\tau=3}^{t} \mathbb{1}_{\left(a_i(\tau-2)=a, a_j(\tau-1)\neq a^*\right)}}{t} & \text{otherwise} \end{array} \right\} \tag{4}$$

In words, players initially see their opponent as being very responsive, but update the perceived responsiveness each period. More precisely, the numerator of $\hat{\gamma}_j(t)$ counts

[18] This assumption is in the spirit of many cognitive hierarchy models in which players view their opponent(s) as being less sophisticated than they themselves are and is also well documented in the literature on overconfidence. See, among others, Roll (1984), Camerer and Lovallo (1999), Benabou and Tirole (2002) and Camerer et al (2004).

[19] Henceforth, all the parameters players perceive about their opponent's behaviour will be denoted with a hat.

the number of times the opponent (j) has *not* best responded to player i's previous action. The denominator simply serves as a normalization factor to ensure that $\hat{\gamma}_j(t)$ lies between 0 and 1. Each time the opponent does not best respond to the previous action, the inertia parameter is reevaluated upward, while each time the opponent does best respond to the previous action, the inertia parameter is reevaluated downwards. We assume, for simplicity that $\hat{\gamma}_j(0) = 0$, and note that the results do not depend materially on the initial value. We also assume that the sophisticated player does not anticipate in period t that he will update γ in period $t+1$ based on the outcome that period, but see Remark 2 for a brief discussion.

Computing current and prospective beliefs of the myopic player, as perceived by the sophisticated player in period t, is a simple matter. In particular, current beliefs are given by:

$$\hat{b}_j^a(t) = \frac{\mathbb{1}_{(a_i(t-1)=a)} + \sum_{s=1}^{t-2} [\hat{\gamma}_j(t)]^s \, \mathbb{1}_{(a_i(s)=a)}}{1 + \sum_{s=1}^{t-2} [\hat{\gamma}_j(t)]^s}, \tag{5}$$

while prospective beliefs at time $u > t$, conditional on the continuation strategy $\sigma^A(t)$ are given by:

$$\hat{b}_j^a(u|\sigma^A(t)) = \frac{1 + \sum_{s=t}^{u-2} [\hat{\gamma}_j(t)]^s + \sum_{s=1}^{t-1} [\hat{\gamma}_j(t)]^s \, \mathbb{1}_{(a_i(s)=a)}}{1 + \sum_{s=1}^{u-2} [\hat{\gamma}_j(t)]^s}. \tag{6}$$

Player i's Beliefs In our experimental sessions, we elicited each player's current period belief, therefore, it remains to detail how players form their prospective beliefs. Given his estimation of his opponent's (current and prospective) beliefs, player i can estimate player j's expected payoff induced by playing action k at time $u \geq t$ given the continuation strategy $\sigma^a(t)$, $a = X, Y$:

$$\hat{E}_j^k(u|\sigma^a(t)) = \sum_{z=X,Y} \hat{b}_j^z(u|\sigma^a(t))\pi_j(k,z) \tag{7}$$

To keep things simple, we will assume that player i believes that his opponent will perfectly best respond in each period to his beliefs. Therefore, player i's prospective beliefs can be written as:

$$b_i^k(u) = \left\{ \begin{array}{ll} 1 & \text{if } k = \arg\max_{k'} \sum_{z=X,Y} \hat{b}_j^z(u|\sigma^a(t))\pi_j(k',z) \\ 0 & \text{if } k \notin \arg\max_{k'} \sum_{z=X,Y} \hat{b}_j^z(u|\sigma^a(t))\pi_j(k',z) \\ 0.5 & \qquad\qquad \text{otherwise} \end{array} \right\} \tag{8}$$

Player i's choice probabilities We assume that player i stochastically best responds given his expected intertemporal payoff $E_i(\sigma^X(t))$. Choice probabilities then take the following logistic form:

$$P_i^X(t) = \frac{\exp\left[\lambda\left[E_i(\sigma^X(t)) - E_i(\sigma^Y(t))\right]\right]}{1 + \exp\left[\lambda\left[E_i(\sigma^X(t)) - E_i(\sigma^Y(t))\right]\right]}. \tag{9}$$

Obviously we have $P_i^Y(t) = 1 - P_i^X(t)$, where $P_i^a(t)$ is player i's propensity to play action a at time t, with λ representing players' sensitivity to expected payoffs. From (9), we can write the likelihood function for individual i at time t as:

$$l_{it} = P_i^X(t)^{\mathbb{1}(a_i(t)=X)} P_i^Y(t)^{\mathbb{1}(a_i(t)=Y)}. \tag{10}$$

Aggregating over individuals and time periods gives us the full likelihood function, which can then be maximized to obtain estimates for λ and δ.

When $\delta = 0$, our model reduces to that of Nyarko and Schotter (2002).[20] We will refer to this as the myopic learning model. When $\delta > 0$, we call it the sophisticated learning model.

Remark 2 (Discussion of the model.) Notice that, while sophisticated players may be far-sighted, they are still not fully rational. In particular, they use an adaptive rule to update the perceived degree of responsiveness of their opponent. Moreover, we have assumed that players do not anticipate in period t that in period $t+1$ they will have an updated γ based on what happened in period t. While one could add perfect foresight into the model it is not clear that this would appreciably affect the qualitative results. With some probability, γ will be updated downwards, making continued teaching more attractive, while with complementary probability, γ will be updated upwards, making continued teaching less attractive. It is not clear whether, in expectation, teaching will be more or less attractive. Depending on which way the expectation leads, the critical value of δ required for teaching to be profitable will likely change, but the essence of the model remains intact.

We have also assumed that the sophisticated player believes his opponent perfectly best responds to her own beliefs. This is made to allow us to focus on the adjustment of γ. The model could be extended to incorporate stochastic best response by the γ-learner, but doing so would greatly complicate the updating procedure since now the sophisticated player would need to jointly update γ and λ in each period. Such a model would be extremely computationally demanding and is beyond the scope of the current paper.

6.3 Results

As stated in the above discussion, our model extends the myopic model with the introduction of sophistication *via* a forward-looking component in players' choice probabilities. This forward-looking component is formalized by the discounted future expected payoffs in the RHS of (3). We estimate the sophisticated learning model for both row and column players. Given that column players had very weak incentive to teach, we conjecture that the sophisticated learning model should not greatly improve the fit above and beyond the myopic model. On the other hand, for row players — particularly when teaching incentives are high — we expect the fit to improve substantially. Our estimation results are reported in Table 7.[21]

We first examine the point estimates of δ. A positive $\hat{\delta}$ indicates that future expected payoffs matter for players and that they take them into account when choosing their current actions. First, look at the results for row players. As can be seen, in three of the four games, we estimate that δ is significantly positive. The one game where we cannot reject the null hypothesis is $\text{TP}_\ell/\text{TC}_h$, and recall that this treatment has the

[20] This follows since we use the elicited stage-game beliefs in (3).

[21] In all our estimations, all standards errors are clustered by individuals.

Table 7 Estimations for each type in each game

Row Players

	Myopic Model				**Sophisticated-learning Model**			
	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$
λ	0.215^{**} (0.086)	0.192^{***} (0.036)	0.555^{***} (0.138)	0.259^{***} (0.043)	0.394^{***} (0.089)	0.224^{***} (0.031)	0.581^{***} (0.108)	0.231^{***} (0.051)
δ	-	-	-	-	0.114^{***} (0.027)	0.187^{***} (0.034)	0.228^{***} (0.046)	0.224 (0.235)
N	340	320	380	300	340	320	380	300
LL	-226.21	-181.32	-215.60	-141.96	-190.12	-151.52	-182.61	-138.35
AIC	454.42	364.64	433.20	285.92	384.22	307.04	369.23	280.71
BIC	458.25	368.41	437.14	289.63	381.89	314.57	377.11	288.12

Column Players

	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$	$\text{TP}_h/\text{TC}_\ell$	TP_h/TC_h	$\text{TP}_\ell/\text{TC}_\ell$	$\text{TP}_\ell/\text{TC}_h$
λ	0.070^{***} (0.019)	0.112^{***} (0.020)	0.096^{***} (0.025)	0.073^{***} (0.013)	0.051^{***} (0.016)	0.112^{***} (0.020)	0.061^{***} (0.018)	0.047^{*} (0.026)
δ	-	-	-	-	0.483 (0.333)	0 (0)	0.569^{**} (0.291)	0.561 (0.726)
N	340	320	380	300	340	320	380	300
LL	-199.73	-151.37	-192.13	-160.17	-194.10	-151.37	-190.32	-159.93
AIC	401.46	304.74	386.26	322.34	392.20	306.74	384.63	323.86
BIC	405.29	308.57	390.30	326.04	399.86	314.28	392.51	331.27

Standard errors in parentheses. Sig. lev.: *: 10%, **: 5%, ***: 1%.

NB: For δ, tests are one-sided tests of $H_0 : \delta = 0$.

weakest teaching incentives. Therefore, except when teaching incentives are weakest, row players incorporate future payoffs into their current decision. Turn now to column players. While δ turns out to be significantly positive in only one game, the point estimates are sizeable. We suspect that the insignificance of δ for column players in most games and for row players in game $\text{TP}_\ell/\text{TC}_h$ comes from the fact that we consistently observe less over responses to X (the actions on which δ is identified) for those players, leading to larger standard deviations for $\hat{\delta}$. Student t-tests show that the estimated δs are not different between the different sub-samples. This is not surprising since players randomly assigned to games and types should not be expected to weight future payoffs in different ways as δ is not payoff-sensitive. Obviously, it does not imply that there should not be differences in teaching behaviour across games. Indeed, for a fixed delta, the stronger (weaker) are the teaching incentives, the more (less) likely we are to observe teaching. We conclude that players do take future payoffs into account when choosing their current actions, although high teaching costs can dominate the expected future payoff gains.

Before examining the other parameters, we pause to discuss whether the addition of the forward looking component (*i.e.,* δ) leads to an improved fit. Of course, since the sophisticated-learning model nests the myopic model, it must be that the log-likelihood weakly increases. To account for the fact that the sophisticated-learning model has an extra parameter, we therefore compare the Akaike and Bayesian Information Criteria (AIC and BIC), which are also reported in Table 7. As can be seen, for row players, in all four treatments, the AIC and BIC are lower in the sophisticated-learning model than in the myopic model, indicating a better fit even after accounting for the extra parameter. For column players, both criteria indicate that the sophisticated-learning model gives a better fit only for the treatment $\text{TP}_h/\text{TC}_\ell$, and the AIC also gives a slight lead to the SL model in treatment $\text{TP}_\ell/\text{TC}_\ell$.

Remark 3 (Mean Squared Deviation.) For each model, we can also calculate the mean squared deviation to get another metric of how the estimated choice probabilities match the data. Such an analysis is problematic because it does not compensate for the fact that the sophisticated learning model contains an additional parameter and as such, must lead to a lower MSD. However, it has the advantage that we can conduct statistical tests. The results are discussed in an online appendix. Stated briefly, the analysis shows that in games where teaching incentives are relatively strong, the MSD for row players are significantly lower when we use our sophisticated-learning model that when we use the myopic model. When teaching incentives are weaker, we can never reject the hypothesis of equality of MSD, this holds for column players across all games and for row players in the game $\text{TP}_\ell/\text{TC}_h$, which had the weakest incentives for teaching. Hence, introducing a forward-looking component into the myopic model helps to track the behaviour of players who are given high incentives to teach and who, according to the our earlier analysis, have the highest propensity to do so.

Turn now to a deeper analysis and comparison of our parameter estimates. First, observe that the estimate of λ for row players is marginally higher in the sophisticated-learning model for the three games where δ is significantly positive, suggesting that at least part of what the myopic model picks up as mistakes are more properly viewed as far-sighted, profit maximizing behaviour. On the other hand, for column players, the exact opposite is found with λ being estimated lower in the sophisticated-learning model.

Next observe that for both the sophisticated-learning model and the myopic model, the estimate of λ is noticeably higher for row players than for column players. Given λ, a belief, b, and teaching cost, θ, the probability a decision maker takes action X can be written as:

$$p(b,\lambda,\theta) = \frac{\exp(\lambda\theta(b-0.8))}{1+\exp(\lambda\theta(b-0.8))}.$$

Recall then from Table 1 that $\theta \in \{5,15\}$ for row players and $\theta = 40$ for column players. Therefore, in order to get the same choice probability for the same belief, we must have λ larger for row players than for column players. Therefore, the overall sensitivity to payoff differences is really $\theta\lambda$. Computing this for the myopic model and we actually have that $\theta\lambda$ is larger for column players than for row players, except in the treatment $\text{TP}_\ell/\text{TC}_h$. For the sophisticated-learning model, the comparison is reversed with $\theta\lambda$ higher for row players, except in treatment $\text{TP}_h/\text{TC}_\ell$, where $\theta\lambda$ remains slightly smaller. Both of these findings reinforce our view that it is row players who are actually teaching. In the myopic model, since row players are teaching, it looks like they are making more mistakes, hence the sensitivity to payoff differences is smaller. However, once long run considerations are accounted for in the sophisticated-learning model, the row players' behaviour no longer looks like mistakes, but instead profit maximization, hence sensitivity increases.

7 Conclusion

In the past decade, several learning models have been devised to describe how people play games. Still, an almost universal assumption in all these models is that players regard their opponents' behaviour as generated by an exogenous process and do not realize that they could influence it *via* their actions. A few recent studies have highlighted the limits of this assumption in various circumstances. More precisely, players might be more sophisticated and attempt to teach their opponent to play a particular action (which is often a Nash equilibrium). This paper has tried to emphasize the determinants of such a strategic behaviour and to show that subjects are often responsive to the incentives that they are given to engage in far-sighted behaviour. Our results have shown the existence of such sophistication in several ways. First, we demonstrated a sophistication bias in players' belief-formation process, which indicates that players take long-run strategic considerations into account. This paves the way for the use of strategic teaching we described in a following step. We found that, particularly when teaching incentives are high (*i.e.,* when teaching is both relatively safe and beneficial), players profitably engage in strategic teaching to drive play to the efficient Nash equilibrium. More precisely, they are much more likely to play their part of the Pareto efficient Nash equilibrium, despite it not being a best response to their static beliefs. Over time, as their opponents learn, this pushes play to the efficient Nash equilibrium. When we made teaching more difficult (by raising the cost of teaching and/or lowering the potential benefit to teaching), we showed that players are much less likely to take long-run considerations into account. In the latter cases, particularly when teaching incentives are the lowest, coordination failure is more prominent.

¿From these results, it is then natural to propose a model that could account for players' behaviour in a unified way across all our games. To address this issue, we extended a model of decision making by introducing a forward-looking component to provide a more accurate description of players using a long-run strategy. Our results showed that this extended model significantly improved the fit for teachers and, by doing so, it also provided a unifying framework to describe players' behaviour in all our games.

References

Battalio R, Samuelson L, van Huyck J (2001) Optimization incentives and coordination failure in laboratory stag hunt games. Econometrica 69:749–764

Benabou R, Tirole J (2002) Self-confidence and personal motivation. Quarterly Journal of Economics 117:871–915

Brandts J, Cooper D (2006) A change would do you good: An experimental study of how to overcome coordination failure in organizations. American Economic Review 96:669–693

Brandts J, Cooper D, Fatas E (2007) Leadership and overcoming coordination failure with asymmetric costs. Experimental Economics 10:269–284

Camerer C, Ho TH (1999) Experienced-weighted attraction learning in normal form games. Econometrica 67(4):827–874

Camerer C, Lovallo D (1999) Overconfidence and excess entry: An experimental approach. American Economic Review 89:306–318

Camerer CF, Ho TH, Chong JK (2002) Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. Journal of Economic Theory 104(1):137–188

Camerer CF, Ho TH, Chong JK (2004) A cognitive hierarchy model of games. Quarterly Journal of Economics 119(3):861–898

Cason T, Lau SHP, Mui VL (2008) Learning, teaching, and turn taking in the repeated assignment game, mimeographed

Chaudhuri A, Schotter A, Sopher B (2009) Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice. Economic Journal 119:91–122

Cheung YW, Friedman D (1997) Individual learning in normal form games: Some laboratory results. Games and Economic Behavior 19(1):46–76

Cho IK, Sargent TJ (2008) Self-confirming equilibrium. In: Durlauf S, Blume L (eds) The New Palgrave Dictionary of Economics

Cho IK, Williams N, Sargent TJ (2002) Escaping Nash inflation. Review of Economic Studies 69(1):1–40

Costa-Gomes MA, Weizsäcker G (2008) Stated beliefs and play in normal form games. Review of Economic Studies 75(3):729–762

Costa-Gomes MA, Crawford VP, Broseta B (2001) Cognition and behavior in normal-form games: An experimental study. Econometrica 69(5):1193–1235

Crawford VP (1995) Adaptive dynamics in coordination games. Econometrica 63(1):103–143

Ehrblatt W, Hyndman K, Özbay E, Schotter A (2008) Convergence: An experimental study of teaching and learning in repeated games, mimeographed

Ellison G (1997) Learning from personal experience: One rational guy and the justification of myopia. Games and Economic Behavior 19:180–210

Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. The American Economic Review 88(4):848–881

Fudenberg D, Levine DK (1989) Reputation and equilibrium selection in games with a patient player. Econometrica 57:759–778

Fudenberg D, Levine DK (1998) The Theory of Learning in Games. The MIT Press

Hopkins E (2002) Two competing models of how people learn in games. Econometrica 70(6):2141–2166

van Huyck J, Battalio R, Beil R (1990) Tacit coordination games, strategic uncertainty and coordination failure. American Economic Review 80(1):234–248

Jehiel P (2001) Limited foresight may force cooperation. Review of Economic Studies 68:369–391

Kandori M, Mailath G, Rob R (1993) Learning, mutation, and long run equilibria in games. Econometrica 61:29–56

Knez M, Camerer C (1994) Creating expectational assets in the laboratory: Coordination in 'weakest-link' games. Strategic Management Journal 15:101–119

Knez M, Camerer C (2000) Increasing cooperation in prisoner's dilemmas by establishing a precedent of efficiency in coordination games. Organizational Behavior and Human Decision Processes 82:194–216

Marcet A, Sargent TJ (1989) Convergence of least squares learning mechanisms in self-referential linear stochastic models. Journal of Economic Theory 48:337–368

Mengel F (2008) Learning by (limited) forward looking players, working Paper RM/08/053

Nyarko Y, Schotter A (2002) An experimental study of belief learning using elicited beliefs. Econometrica 70(3):971–1005

Offerman T, Sonnemans J (2001) Is the quadratic scoring rule behaviorally incentive compatible?, mimeographed

Offerman T, Sonnemans J, Schram A (2001) Expectation formation in step-level public good games. Economic Inquiry 39:250–269

Roll R (1984) Orange juice and weather. American Economic Review 74(5):861–80

Rutström EE, Wilcox NT (2004) Learning and belief elicitation: Observer effects. Working paper

Samuelson L (1998) Evolutionary Games and Equilibrium Selection. The MIT Press

Stahl DO, Wilson PW (1995) On players models of other players: Theory and experimental evidence. Games and Economic Behavior 10(1):218–254

Terracol A, Vaksmann J (2009) Dumbing down rational players: Learning and teaching in an experimental game. Journal of Economic Behavior and Organization 70(1-2):54–71

Weibull JW (1997) Evolutionary Game Theory. The MIT Press

Zeiliger R (2000) A presentation of regate, internet based software for experimental economics., http://www.gate.cnrs.fr/∼zeiliger/regate/regateintro.ppt, Lyon: GATE

Appendix A. Instructions (translated from French)

Thank you for participating in this experimental session. During this session, upon the choices you make, you may be able to earn a significant amount of money which will be paid you in private at the end of the experiment. Your identity and those of the other participants will never be disclosed.

This session contains **20 repetitions** (which will be labelled "rounds" on your screen). Your final payment corresponds to the sum of the payoffs you earn at each repetition. More precisely, during the 20 rounds of this session, you will make points labelled in Unités Monétaires Expérimentales (UME). At the end of this session, your total payment in UME will be converted into Euros at the rate:

100 UME = €2.1

During this session, you will **not** be allowed to communicate with other participants. If you have any questions, please raise the hand and the experimenter will publicly answer.

Type and matching

At the beginning of the session, you will be attached a "type", you can be either of type 1 or of type 2. **Your type will remain the same for the whole session**. Moreover, you will be matched with a pair partner, picked up at random at the beginning of the session **among the participants whose type is different from yours**. For example, if you are of type 1 (resp. type 2), your pair partner will be of type 2 (resp. type 1). **Your pair partner will be the same for the whole session**.

Your decisions

In each round, every participant can choose among **2 decisions: X or Y**. The payoff associated to your decision in a given round depends on your own decision and the decision of your pair partner. These payoffs are presented in Tables B.1 or B.2 below if you are respectively of type 1 or of type 2.

Prediction of other people's decisions

Prior to choosing a decision in each round, you will be given the opportunity to earn additional money by predicting the decision your pair partner will take in the current round. Thus, at the beginning of each round, you will be asked the following two questions:

- On a scale from 0 to 100, how likely do you think your pair partner will take decision X?
- On a scale from 0 to 100, how likely do you think your pair partner will take decision Y?

For each question you have to key in a number greater than or equal to 0. The sum of the two numbers you enter has to equal 100. For example, suppose that you think that there is a 65% chance that your pair partner will take decision X and a 35% chance that your pair partner will take decision Y. In this case, you will key in 65 in the upper box on the screen and 35 in the other box. At the end of each round, we will look at the decision actually made by your pair partner and compare his decision to your prediction. We will then pay you for your predictions as follows. Consider the above example: you entered 65% for decision X and 35% for decision Y. Suppose now that your pair partner actually chooses Y. In this case, your payoff for your predictions will be:

$$4[2-(1-0.35)^2-(0.65)^2].$$

In other words, you will be given a fixed amount of $4 \times 2 = 8$ points (in UME) from which we will subtract an amount which depends on how inaccurate your predictions were. To do this, when we find out what decision your pair partner has made, we will take the number you assigned to that decision, in this example 35% (or 0.35) on Y, subtract it from 100% (or 1), square it and multiply by 4. Next, we will take the numbers assigned to the decision not made by your pair partner, in this case the 65% (or 0.65) you assigned to X, square them and multiply by 4. These two squared numbers will then be subtracted from the 8 points we initially gave you to determine the final payoff associated to your predictions for the current round.

Note that since your predictions are made before you know what your pair partner has actually chosen, the best thing you can do to maximize the expected size of your prediction payoff is to simply state your true beliefs about what you think your pair partner will do. Any other predictions will decrease the amount you can expect to earn as a prediction payoff. Note also that you can not lose points from making predictions but can only earn more points. The worst you can do is predicting that your pair partner will take one particular decision with 100% certainty but it turns out that he actually takes a different decision. In this case, you will earn 0 point. Similarly, the best you can do is to guess correctly and assign 100% to that decision which turns out to be the actual decision chosen. Here, you will keep the whole 8 points amount that was given to you at the beginning of the current round.

In each round, you will have two minutes to enter a correct report. If you make a mistake in a report, i.e. if you enter two numbers which sum is different from 100, or if your report is incomplete, you will be able to retry as many times as you want subject to the fact that you have enough time left to do so. If the available time runs out while you have not entered a correct report, the game continues and you will take your decision but you will not get any payoff for your predictions at the current round.

The computer screens

In each round, you will enter your predictions and take your decisions on different screens represented below.

In the first screen, you will have to report your predictions. You have to enter one number for each decision in the box next to the corresponding question. You will see on your screen the time remaining to report your predictions in figures at the upper right of the screen. Below the two questions, you have a calculator that automatically provides you, in the box "**Sum**", the sum of the numbers you enter and show you, in the box "**Rest**", 100 minus the number you have already entered so that it would made computations easier while reporting your predictions. You can change your report at any time provided that you have any time left to do so and when your report sounds to you satisfactory, click **OK** to proceed to the next screen to take your decision.

In the next screen, you will have to pick up a decision among the two decisions available. To do so, you have to click on the box corresponding to your choice. You have as much time as you want to take your decision.

Once you have reported your predictions and taken your decision, you will get information about the current round. More precisely, you will see recapitulated on a final screen your decision, the decision of your pair partner, your payoff and the payoff of your pair partner associated to your decisions along with the predictions you reported and your prediction payoff. Your predictions, your decisions, the decisions of your partner, and your respective decision payoffs will remain present during the whole session on the bottom of your screen in the table which recapitulates the history of the game by round, so that you will always be able to track what happened in previous round and you will always see which round you are in. Moreover, the last line in the table reminds you of your type so that you could always look at the payoff tables in an appropriate way.

Your Final Payment

The payoff associated to your predictions will be in addition of what you will make with your decisions. Your final payment in UME will simply be **the sum of all payoffs** you will make throughout the 20 rounds of this session; it is this total payoff that will be converted into Euros at the above rate.

Tables and Figures From Experimental Instructions

The two following Tables show subjects' payoffs for game $\text{TP}_h/\text{TC}_\ell$, apart from the payoffs contained in these Tables, the instructions are obviously the same in all games.

Table B.1. Description of Payoffs for Type 1 Players

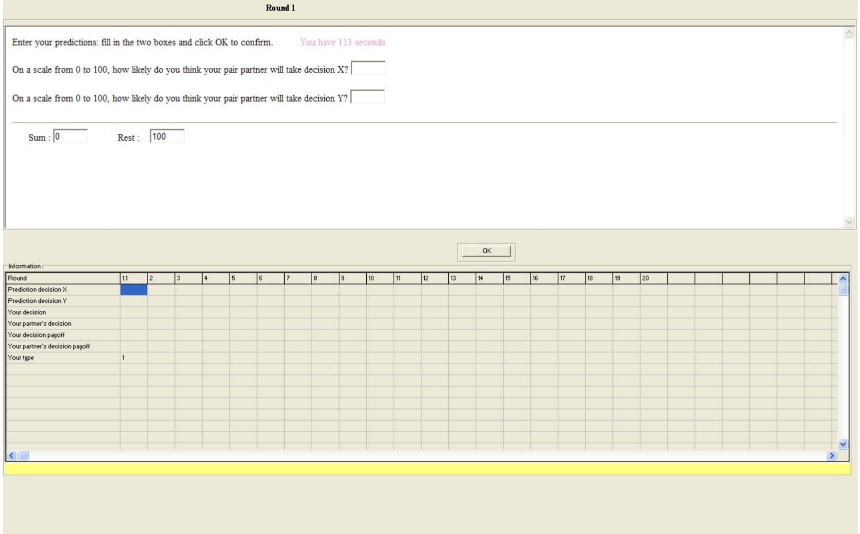
Your decision	Decision of your pair partner	Your payoff	The payoff of your pair partner
X	X	40	45
X	Y	8	37
Y	X	39	0
Y	Y	12	32

Table B.2. Description of Payoffs for Type 2 Players

Your decision	Decision of your pair partner	Your payoff	The payoff of your pair partner
X	X	45	40
X	Y	0	39
Y	X	37	8
Y	Y	32	12

The screen where you will be asked to report your predictions (the last line of the table shows the appropriate type, type 1 in this example)

Fig. 2 Screenshot 1



Round 1

Enter your predictions: fill in the two boxes and click OK to confirm. You have 113 seconds

On a scale from 0 to 100, how likely do you think your pair partner will take decision X?

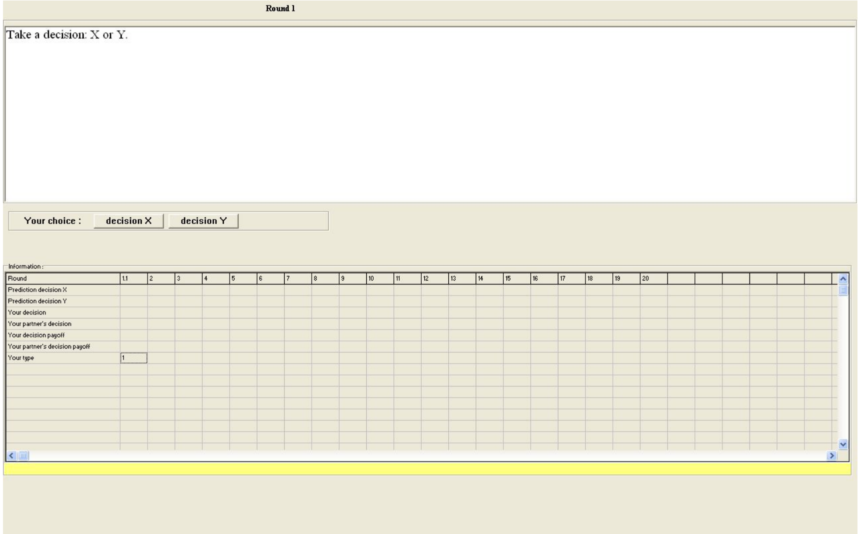
On a scale from 0 to 100, how likely do you think your pair partner will take decision Y?

Sum : 0 Rest : 100

Round	1.1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Prediction decision X																				
Prediction decision Y																				
Your decision																				
Your partner's decision																				
Your decision payoff																				
Your partner's decision payoff																				
Your type	1																			

The screen where you will be asked to take your decision (the last line of the table shows the appropriate type, type 1 in this example)

Fig. 3 Screenshot 2



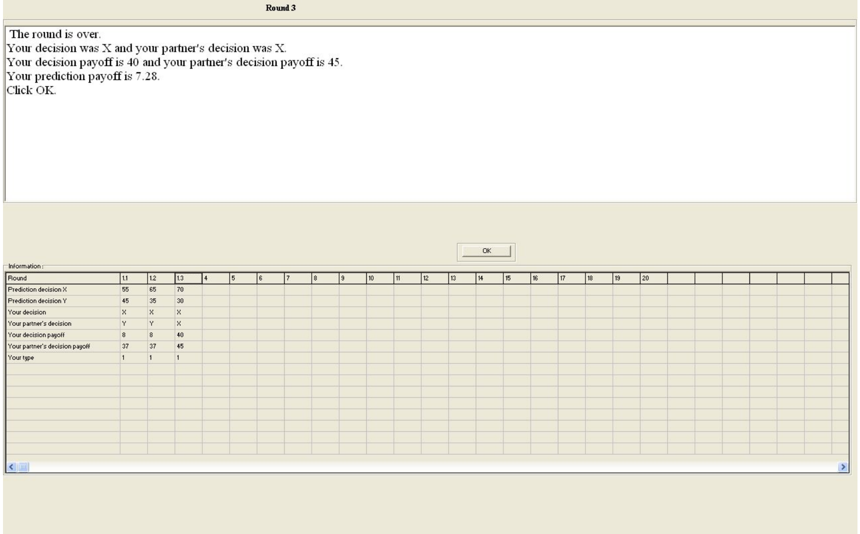
Round 1

Take a decision: X or Y.

Round	1.1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20						
Prediction decision X																										
Prediction decision Y																										
Your decision																										
Your partner's decision																										
Your decision payoff																										
Your partner's decision payoff																										
Your type	1																									

The following screen was not contained in the instructions given to the subjects. It shows an example of how information were recapitulated at the end of each round.

Fig. 4 Screenshot 3



The round is over.
Your decision was X and your partner's decision was X.
Your decision payoff is 40 and your partner's decision payoff is 45.
Your prediction payoff is 7.28.
Click OK.

Round	1.1	1.2	1.3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Prediction decision X	55	65	70																	
Prediction decision Y	45	35	30																	
Your decision	X	X	X																	
Your partner's decision	Y	Y	X																	
Your decision payoff	8	8	40																	
Your partner's decision payoff	37	37	45																	
Your type	1	1	1																	