Learning in Network Games*

Jaromír Kovářík[†]
University of the Basque Country

Friederike Mengel [‡§]
University of Nottingham and
Maastricht University

José Gabriel Romero[¶]
University of Santiago de Chile

October 1, 2012

Abstract

We report the findings of an experiment designed to study how people learn and make decisions in network games. Participants in our experiment interact in an (Anti-)Coordination game for 20 rounds with their neighbours in a network. Our experimental design enables us to observe *both* which actions participants choose and which information they consult before making their choices. We use this information to estimate learning types using maximum likelihood methods. There is substantial heterogeneity in learning types. However, the vast majority of our participants are categorized either as reinforcement learners or (myopic) best-response learners. Network topology and player position in the network have limited influence on the estimated distribution of learning types. We do, however, find some differences. In particular, players in networks with cycles and players in positions with more neighbors tend to be characterized by simpler learning rules. Our results suggest that, while broad categories of learning are stable across contexts, players adjust towards simpler learning rules in more complex environments.

JEL Classification: C72, C90, C91, D85.

Keywords: Experiments, Game Theory, Heterogeneity, Learning, Maximum Likelihood Method, Networks.

---

*We wish to thank Clayton Featherstone, Sanjeev Goyal, Matthew Jackson, Muriel Niederle, Aljaz Ule and Marco van der Leij as well as audiences in Stanford, Navarra, Cambridge, USACH, Universidad Católica de Chile, Shanghai (ESWC 2010), Copenhagen (ESA 2010), Oslo (ESEM 2011) and Santiago (LAMES 2011) for valuable comments. Friederike Mengel thanks the Dutch Science Foundation (NWO, VENI grant 016.125.040) and Jaromír Kovářík the Basque Government (IT-223-07) and the Spanish Ministry of Science and Innovation (ECO2009-09120) for financial support.

†Dpto Fundamentos Análisis Económico I & Bridge, Universidad del País Vasco/Euskal Herriko Unibersitatea, Av. Lehendakari Aguirre 83, 48015 Bilbao, Spain (jaromir.kovarik@ehu.es).

‡School of Economics, University of Nottingham, University Park Campus, NG7 2RD Nottingham, United Kingdom (friederike.mengel@nottingham.ac.uk).

§Department of Economics (AE1), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.

¶Departamento de Economia, USACH, Alameda 3363, Santiago, Chile (gabriel.romero@usach.cl).

1 Introduction

Most scholars now agree that economic agents arrive at their decisions in strategic situations via a process of learning and economists have developed a variety of theoretical models of learning in games. These models, however, often lead to very different predictions both in the short- and in the long-run (Fudenberg and Levine, 1998). As making predictions is one of the ultimate goals of economics, it is important to understand how economic agents learn, whether there is heterogeneity, and to which extent learning rules employed by agents remain stable across different contexts. Abstracting from these issues may lead to imprecise predictions of behavior at the individual and aggregate level and this may in turn impede correct evaluations of final impacts of policy interventions and economic shocks.

A considerable amount of research has been conducted to understand how people learn in games. This research has provided mixed evidence so far. Models that have found support in some studies have been rejected in others.[1] By far the most common approach to study learning in experiments has been the representative-agent model. Under this approach a single learning model is estimated to explain the average or median behavior of participants. One downside of the representative-agent approach is that, if there is heterogeneity in learning types, it is far from clear how robust the insights are to small changes in the distribution of types or whether comparative statics predictions based on the representative agent will be correct (e.g. Kirman, 1992). In addition, Wilcox (2006) shows that in the presence of heterogeneity representative agent models can produce significant biases favoring reinforcement learning relative to belief learning models (see also Cheung and Friedman, 1997, or Ho et al., 2008). Two exceptions to this literature are Camerer and Ho (2002), who assume that agents fall into two segments of subpopulations with different parameter values for each, and Camerer et al. (2002), who estimate a mixture of standard and sophisticated EWA learners in the population.

Another approach has been to estimate learning models individually for each subject (Camerer, Ho, and Wang, 1999; Cheung and Friedman, 1997; Ho et al., 2008). However this approach is likely to lead to small-sample biases (Cabrales and Garcia Fontes, 2000; Wilcox 2005) and estimations are only consistent if the experiment involves "sufficient" time periods, where "sufficient" can often mean practically infeasible in a typical experiment.[2]

In addition, many studies restrict the information feedback given to participants thereby ruling out some learning models ex ante. If e.g. no information about payoffs of other participants is provided then payoff-based imitation learning is not possible. But even if information feedback is extensive, estimation of learning types solely on basis of observed behavior may easily fail to detect the underlying data-generating processes (Salmon, 2001). Hence, while the existing literature has provided valuable insights into how people learn in specific situations, most studies were not designed to address the questions of individual heterogeneity and stability across contexts we mention above.

In this paper we attempt to address these issues. We conduct an experiment where information about past play and payoffs of all participants as well as about network neighbors is available, but where we keep track of which information each participant consults between rounds. We combine this information with information about observed choices to estimate a distribution of learning types using maximum likelihood methods. The advantage of observing both behavior and information is that even if different learning rules predict the same choices at the same decision node, they can be distinguished as long as different information is needed in each case to identify the correct choice.

We study strategic interactions in networks. In networks (compared to random matching or fixed

[1]See Camerer (2003, Chapter 6), Erev and Roth (1998), Mookherjee and Sopher (1997), Kirchkamp and Nagel (2007) or Feltovich (2000) among others.

[2]Cabrales and García-Fontes (2000, Footnote 17) report that the precision of the estimates starts to be "reasonable" after observing around 500 periods of play.

pairwise matching scenarios) it is more often possible to distinguish learning models via information requests. For example, consider myopic best response and forward-looking learning. Under random matching an agent needs to know the distribution of play in the previous period irrespective of whether she is learning via myopic best responses or whether she is forward-looking. In a network, though, a myopic best responder needs to know *only* the past behavior of her first-order neighbors (who she interacts with), while a forward-looking learner may need to know the behavior of her second-order neighbors to be able to predict what her first-order neighbors will choose in the following period.[3] An additional advantage of this design is that it allows us to systematically change the network topology (moving e.g. from very homogeneous to heterogeneous situations) and see how this affects the estimated distribution of learning types. We can also ask whether an agent's position *within* a network (e.g. central vs. peripheral) affects the way she learns. Hence our study allows us to address two key questions that most previous studies have found difficult to address: the question of individual heterogeneity and the question of how stable learning is across contexts.

Participants in our experiment interacted in a 4×4 Anti-Coordination Game. We hoped that with 4×4 games we would eventually get convergence to Nash equilibrium, but that convergence would not be immediate. Slow enough convergence is necessary to be able to study learning across a number of rounds. Compared to pure Coordination games, Anti-Coordination games also have the advantage that different learning rules prescribe different play more often and, compared to e.g. conflict games, they have the advantage that standard learning models do converge.

In our analysis we apply a methodology first introduced by El-Gamal and Grether (1995) and extended by Costa-Gomes et al. (2001; CCB, henceforth). CCB monitor subjects' information look-ups and use an error-rate analysis to develop a procedural model of decision making, in which a subject's type first determines her information look-ups, possibly with error, and her type and look-ups jointly determine her decision, again possibly with error. In our model, each player learns according to a certain learning rule, which is drawn from a common prior distribution. We consider four prominent learning models as possible descriptions of subjects' behavior. One of our learning types is reinforcement learning, another is imitation learning, and two models correspond to belief-based learning (myopic best response and forward-looking learning). A potential downside of our approach is that the fact that participants have to request information may *per se* distort their decisions. We try to understand whether this is the case by comparing decisions in the main treatments with additional treatments.

In total our experiment consists of six treatments. Three treatments with endogenous information search (for three different network topologies) and three control treatments with the same networks but *without* endogenous information search. In these full information treatments participants are given all the information that can be requested in the former treatments by default. We use these control treatments to see whether the possibility of search *per se* affects behavior and whether participants request all the information they would naturally use in making their decisions. We find no significant differences in behavior between the control treatments and the treatments with endogenous information search.

We now briefly summarize our main results. There is substantial heterogeneity in the way people learn in our data. However, most agents can be classified as either reinforcement learners or belief learners. Even though we observe significant effects of the network topology on subjects' behavior (which actions they choose), our results suggest that networks have limited influence on how people learn. In fact the estimates are virtually identical in two of our networks. We do find, however, that in our (star-like) network with the highest variance in the degree distribution all participants are

[3] Which information she needs exactly will of course depend on her theory about how her first-order neighbours learn. However, it is clear, that a myopic best response learner does *not* need information beyond her first-order neighbourhood.

best characterized by belief learning rules, while there is no evidence for reinforcement learning.

On the other hand, player position within the network seems to affect learning. Peripheral players (with fewer links) are best characterized by myopic best response as opposed to central players who are characterized by a mixture of reinforcement learners and myopic best responders. This difference is intuitive. Forming beliefs about opponents' possible actions (as in myopic best response) requires more reasoning resources and hence may become more costly in terms of reasoning cost if an agent has more neighbors. Moreover, as most players in our star-like network only have one neighbor this can also explain why we only observe belief learning in this network. Hence, people seem to be shifting towards simpler and cognitively less demanding rules in more complex situations (being it more neighbors or a more complex network architecture).

Because almost all participants can be described by either reinforcement learning or (myopic) belief-based rules, our results support the assumptions of EWA (Camerer and Ho, 1998, Camerer et al., 2002). EWA includes reinforcement and belief learning as special cases as well as some hybrid versions of the two. Unlike in EWA we do not restrict to those models ex ante, but our results suggest that - at least in the context considered - a researcher may not be missing out on too much by focusing on those models. However, while EWA should be a good description of behavior at the aggregate level, at the individual level only about 16% of our participants request information consistent with *both* reinforcement learning and belief-based learning rules.

To assess how important it is to use information beyond subjects' actions, we compare the estimated population shares with estimations where we disregard information searches. We detect large biases in these estimates. Estimations based solely on observed action choices lead us to accept certain learning rules that subjects could not have been using, simply because they did not consult the minimum amount of information necessary to identify the corresponding actions. Since we use a relatively large 4×4 game, which allows to distinguish learning rules more easily on the basis of behavior only, this problem is likely to be more severe in smaller 2×2 games often studied in experiments.

The paper proceeds as follows. Section 2 describes in detail the experimental design. Section 3 gives an overview of behavior using simple descriptive statistics. Section 4 introduces the learning models. Section 5 contains the econometric framework and our main results. Section 6 presents additional results and Section 7 concludes. Some additional tables and the experimental Instructions can be found in Appendix.

2 Experimental Design

2.1 General Setup

In all treatments in our experiment participants repeatedly played the symmetric two player game depicted in Table 1 with their (first-order) neighbors in the network. Within each session the networks were fixed, which means that each participant played with the same neighbors in all of 20 periods. Each player had to choose the *same* action against all her neighbors. If participants were allowed to choose different actions for their different neighbors, the network would become irrelevant for choices and many learning rules would become indistinguishable in terms of information requirements.

Payoffs in each round are given by the average payoff obtained in all the games against the neighbors. We chose to pay the average rather than total payoffs to prevent too high inequality in earnings due to different connectivity. The game payoffs are expressed in terms of Experimental Currency Units (ECU), which were converted into Euros at the end of the experiment at exchange rate 1 Euro to 75 ECU. The 4×4 one-shot game involves elements of both coordination and anti-

	A	B	C	D
A	$20, 20$	$\mathbf{40, 70}$	$10, 60$	$20, 30$
B	$\mathbf{70, 40}$	$10, 10$	$30, 30$	$10, 30$
C	$60, 10$	$30, 30$	$10, 10$	$\mathbf{30, 40}$
D	$30, 20$	$30, 10$	$\mathbf{40, 30}$	$20, 20$

Table 1: The Game

coordination. Equilibria (A,B) and (B,A) are efficient, while there is a sense in which actions C and D are less risky. C maximizes payoffs if neighbors choose uniformly at random and D is the maxmin choice. Hence there is an element of Coordination on either efficient or risk-dominant behavior. On the other hand, there is also an element of Anti-Coordination, since within each subset $\{A, B\} \times \{A, B\}$ and $\{C, D\} \times \{C, D\}$ each player has an equilibrium which she strongly prefers and all Nash equilibria are such that the two players have to choose different actions. We chose a 4×4 rather than a 2×2 game, because (i) we hoped that this would generate sufficiently slow convergence to equilibrium to be able to analyze learning in a meaningful way and (ii) a larger game makes it already easier to identify a larger number of different learning rules from observing agents' choices only.

The treatments differed along two dimensions: network architecture and information accessibility. Throughout the paper we denote network architectures with numbers 1, 2 and 3 (see Figures 1-3) and information levels with capital letters N (endogenous) and F (full information). In Subsection 2.2 we discuss our three network topologies and in Subsection 2.3 we explain the information conditions.

2.2 Network Topology

Figures 1-3 present the three network architectures used in the experiment and Table 2 summarizes the most standard network characteristics of these networks.[4] The three networks are very similar in terms of most network characteristics, with the exception of degree heterogeneity, measured by the variance in degree, $\sigma^2(\kappa) = \sum_{i=1}^{8}(\kappa_i - \overline{\kappa})^2$, where κ_i is the degree of agent i (i.e. the number of i's first-order neighbors) and $\overline{\kappa}$ the average degree in the network. We chose the networks in such a way that starting from the homogeneous network, the circle, heterogeneity in degree is varied while the other network characteristics are kept approximately constant. We selected networks with different variances $\sigma^2(\kappa)$ to see whether and how learning is affected if there are strong asymmetries in the environment. Note also that two of our networks contain cycles, while the third one does not.

An equilibrium in a network game (in our experiment of 8 players) is obtained when all players choose an action that is a best response to whatever their neighbors choose. All our networks are designed such that many pure strategy equilibria exist in the one-shot network game. A table describing all strict Nash equilibria in the three networks can be found in the Appendix. Coordinating a network of 8 players on any one of the many possible equilibria is possible, but not obvious. We expected to see mis-coordination in early periods, but hoped to see learning and convergence to equilibrium afterwards.

[4]Obviously, one could define many more network characteristics. In Table 2 we present measures that can be found in standard textbooks; e.g. Vega-Redondo (2007).

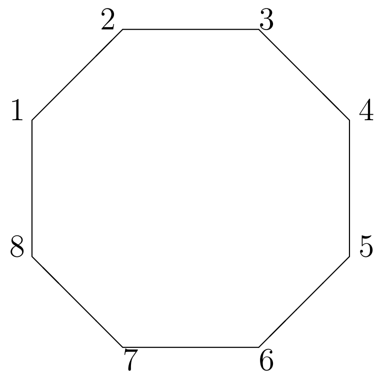


Figure 1: Treatments N-1 and F-1

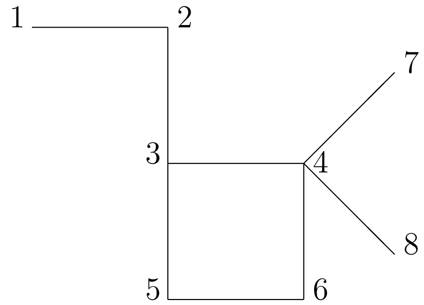


Figure 2: Treatments N-2 and F-2

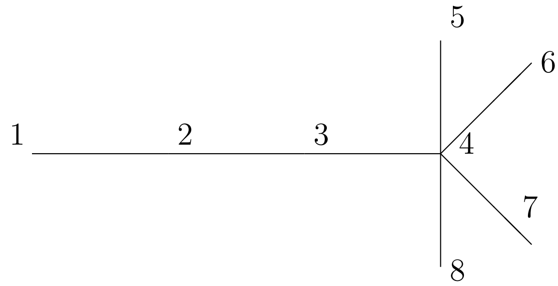


Figure 3: Treatments N-3 and F-3

Of course, each of these networks also allow for many Nash equilibria of the repeated (20 period) game. Since our focus here is on learning we will not discuss or compare these equilibria any further. However, we can say at this stage that in *none* of the networks (in any of the treatments) behavior corresponded even approximately to a Nash equilibrium of the repeated game. Behavior did converge, though, to a Nash equilibrium of the one-shot game in several networks (see below).

2.3 Information

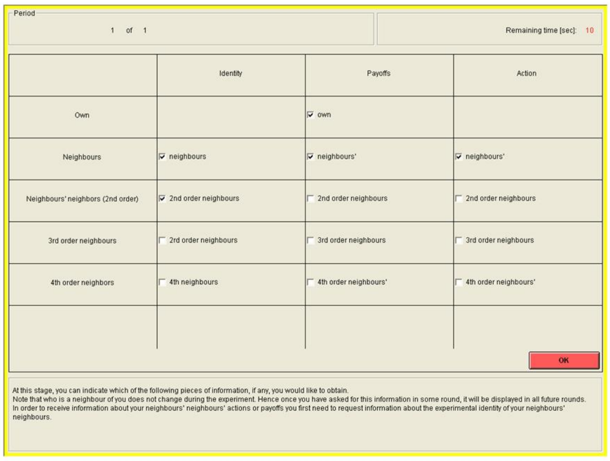
Another treatment variable concerns information about histories of play. The benchmark cases are provided by treatments $N-1$, $N-2$ and $N-3$. In these *endogenous information* treatments, we did not provide our participants with any information by default. Rather, at the beginning of each round they were asked which information they would like to consult. They could ask for three types of information: (i) the network structure, (ii) past action choices and (iii) past payoffs of their network

	$N-1$	$N-2$	$N-3$
Number of players	8	8	8
Number of links	8	8	7
Average degree $\overline{\kappa}$	2	2	1.75
$\sigma^2(\kappa)$	0	8	16.5
Charact. path length	2.14	2.21	2.21
Clustering coeff.	0	0	0
Average betweenness	0.42	0.40	0.37
Variance betweenness	0	0.21	0.21

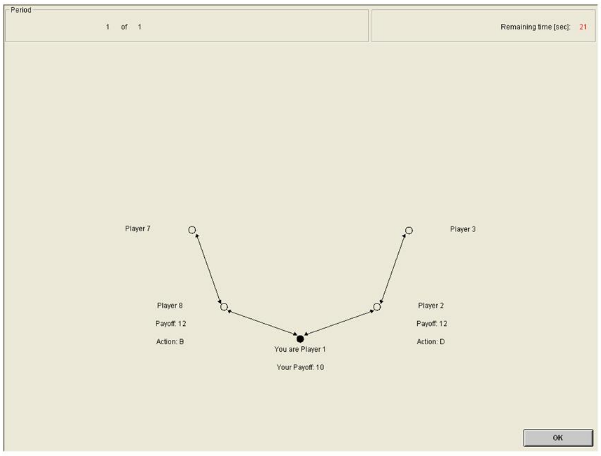
Table 2: Network Characteristics.

neighbors. More precisely, if a subject asks for the network position of her first order neighbors she is shown how many neighbors she has and their experimental identity (which is a number between 1 and 8; see Figures 1-3). With second order neighbors, she is shown their experimental identity as well as the links between the first and second-order neighbors. The same holds for third and fourth-order neighbors. Regarding actions and payoffs, subjects were shown the actions and/or payoffs of their first-, second-, third- and/or fourth-order neighbors if they asked for this information. Participants were also not shown their own payoff by default, but instead had to request it. This design feature allows us to have complete control over which information people hold at any time of the experiment.

We placed two natural restrictions on information search. First, subjects were only allowed to ask for the actions and/or payoffs of subjects whose experimental identity they had previously requested. Second, they were not allowed to request the experimental identity of higher order neighbors without knowing the identity of lower order neighbors. Figures 4(a) - 4(b) show both the screen, in which each subject had to choose the type of information she desired, and how the information was displayed after subjects had asked for it. Each piece of information about actions and/or payoffs had a cost of 1 ECU. Requesting information about the network had a larger cost of 10 ECU, since, once requested, this information was permanently displayed to the participants.



(a) Screen: Information Requests



(b) Screen: Information Display

Imposing a (small) cost on information search is a crucial element of our design. Of course, even though costs are "small" this does affect incentives. We imposed costs to avoid that participants request information they are not using to make their decisions. We also conducted one treatment that coincided with treatment $N-2$ but where there was no cost at all to obtaining information. In this treatment behavior did not differ significantly from $N-2$, but participants asked for all the information (almost) all the time. This means that without costs monitoring information search does not help us to identify learning rules.[5]

To see whether information search per se affects behavior (e.g. because participants do not look up "enough" information due to the costs) we conducted three control treatments with full information. In those treatments $F-1$, $F-2$ and $F-3$ there was no information request stage and all the information was displayed at the end of each period to all participants. We call these the *full information* treatments. We can use those treatments to study whether convergence is slower or whether there are any other differences in behavior induced by the existence of costly information search. We did not find significant differences between the F and N treatments (see below). Table 3 summarizes the treatment structure of the experiment.[6]

	Network 1	Network 2	Network 3
Endogenous Information (N)	40 (800; 5)	56 (1120; 7)	40 (800; 5)
Full Information (F)	24 (480; 3)	24 (480; 3)	24 (480; 3)
Total	64 (1280; 8)	80 (1600; 10)	64 (1280; 8)

Table 3: Treatments and Number of Subjects (Number of Observations; Number of Independent Observations).

The experiment was conducted between May and December 2009 at Maastricht University using the software Z-tree (Fischbacher, 2007). A total of 224 students participated. The experiment lasted between 60-90 minutes. Each 75 ECU were worth 1 Euro and participants earned between 7,70 and 16,90 Euros, with an average of 11,40 Euro.

3 Descriptive Statistics

In this section, we provide a brief overview of the experimental findings, regarding convergence to equilibrium and information searches of subjects. Since we want to focus on estimation of learning types in this paper we refer readers to Kovarik et al. (2011) for a detailed analysis of the observed behavior in this experiment.

3.1 Behavior

Network topology has a strong impact on behavior. More precisely, there is an non-monotonic effect of degree heterogeneity on convergence to equilibrium. We observe relatively high convergence in Network 2, intermediate levels in Network 1, followed by Network 3. The entire network converges to an equilibrium 12% (28% and 0%) of cases in $N-1$ ($N-2$ and $N-3$, respectively). All the differences are statistically significant if we take each network (rather than individual) as an

[5]As alternative approach was taken by CCB. They use the computer interface MouseLab. However, as they state "the space of possible look-up sequences is enormous, and our subjects' sequences are very noisy and highly heterogeneous" (p. 1209). As a result, they make several assumptions to be able to work with the data. We avoid some of these problems with our design.

[6]The table does not contain the treatment $N-2$ without costs mentioned above. We will not discuss this treatment any further, but results are of course available upon request. Other than the treatments reported we didn't conduct any other treatments or sessions and we did not run any "pilot studies".

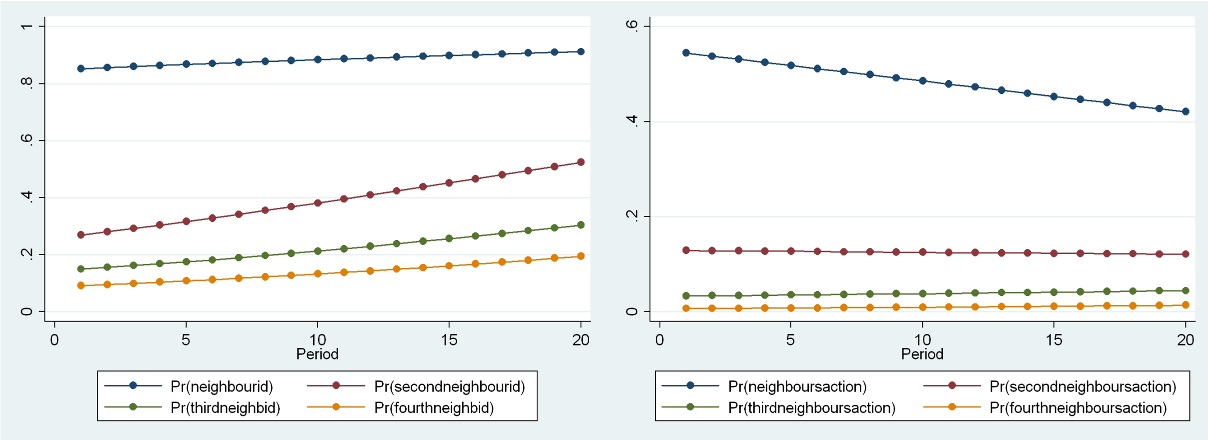


Figure 4: Information Search. *Left:* Identity of First-, Second-, Third- and Fourth-Order Neighbors. *Right*: Action Choices of First-, Second-, Third- and Fourth-Order Neighbors. Note the different scales of the y-axis.

independent observation (Mann-Whitney, $p < 0.01$). Behavior is not statistically different, however, along the information dimension (convergence rates are 13%, 46% and 13% for F-1, F-2 and F-3 respectively). In all treatments participants eventually coordinated on a Nash equilibrium where all players choose either C or D (if they coordinate at all). (A table containing all the strict Nash equilibria of the one-shot network game can be found in Appendix A.) Hence different information conditions (i.e. full vs. endogenous information) do *not* seem to change the behavior of experimental subjects.

3.2 Information

Given the important role information requests will play in learning type detection in later sections, we describe the search patterns in more detail.

Network Structure. Figure 4 illustrates the information searches concerning the network structure, aggregated over treatments. In the first round 77.5%, 76.8% and 72.5% of subjects in $N-1$, $N-2$ and $N-3$, respectively, asked for the identity of direct neighbors in the first round. Roughly 90% of individuals end up demanding this information (92.5, 89.3 and 87.5%, respectively) by the last round of the experiment. Around 45% of subjects request the network structure up to their second-order neighbors (35%, 50% and 50% for $N-1$, $N-2$ and $N-3$, respectively) by the last round. Only 12.5%, 23.2% and 12.5%, respectively, request information about the entire social network. Remember that information about the network structure - once requested - was permanently displayed.

Payoffs. Slightly less than 50% of individuals ask for their own payoff to be displayed. Only about 11.9%, 10.7% and 11.6% for $N-1$, $N-2$ and $N-3$, respectively, require information about the payoffs of their first-order neighbors. These percentages drop statistically to zero for more distant individuals.

Actions. Around 50% of individuals pay to learn past actions of their opponents (i.e. their direct neighbors in the network). There is no statistical difference across the three networks. The percentages decline over time, which we attribute to convergence to equilibrium. Despite the strategic effect of second-order neighbors' actions on the play of direct opponents, the interest in their behavior is relatively small. After period 3 in all N-treatments, only around 18% of individuals request past action choices of agents at distance two. And only a negligible fraction of subjects look for information about choices of third- or fourth-order neighbors. Figure 4 shows the evolution of the action-related information requests over time, aggregated over all the treatments.

4 Framework

This section presents the framework we use to estimate the learning types of our subjects. Our fundamental view is that subjects learn about stage-game strategies rather than playing repeated-game strategies from the outset. Each subject's behavior is then determined, with error, by one of four learning rules (types). The type of each agent is determined by independent draws from the same distribution and remains constant over the 20 rounds she plays. Our goal is to estimate the distribution of types.

We consider four possible learning types. One rule is reinforcement, another rule is based on imitation, and two rules are belief-based. The criterion for the selection of these learning types was their prominent role in the theoretical and experimental literature. In what follows, we describe each of them informally; the exact algorithms used for each learning model considered in this paper can be found in Appendix:

1. Under *Reinforcement learning* (*RL*) participants randomize between actions with probabilities that are (linearly) proportional to past payoffs obtained with these actions.
2. *Payoff-based imitation* (*PBI*) selects the most successful action from the previous period (i.e. the action with the highest average payoff) in an agent's first-order neighborhood including the agent herself.
3. Under *Myopic best responses* (*MBR*) players choose a myopic best-response to the last period play of their first-order neighbors.
4. *Forward-looking (FL)* subjects assume that their first-order neighbors are myopic best responders and best-respond to what they anticipate their first-order neighbors to play in the following period.

All the above learning rules are adaptive. The forward-looking rule is probably the most sophisticated among them and closest to capturing aspects of repeated interaction. We exclude hybrid models, such as experience-weighted attraction of Camerer and Ho (1999). However, we can say something about how well EWA will be able to describe behavior by looking at how well its component rules perform. The reader may also wonder why we did not include level-k learning rules or similar. There are multiple reasons for this. The main reason is that level-k learning - despite its name - is not defined as an explicitly dynamic learning model. As a consequence it is not clear which information a level-k learner should request or how they should update their beliefs about the distribution of k in the population upon receiving new information. Section 6 reports estimations with some variants of the above learning rules.

4.1 Identifying Learning Rules from Decisions

Obviously, it is only possible to identify learning rules if different rules predict different actions and/or information requests in the experiment. Concerning action choices, for every period and subject we compute which action she should have played according to each of the learning rules.

Table 4 presents the average number of rounds (out of 19) in which - given the behavior in the experiment - two different learning types predict different action for the participants.

In all treatments, the number of rounds RL prescribes different behavior than our imitation learning model ranges from 11 to 15 rounds. Reinforcement is also well separated from belief-based learning models (in at least 7/19 rounds). The number of rounds in which PBI's decisions are different from those of belief learning ranges from 11 to 15 rounds. Finally, the number of rounds in which MBR and FL's decisions differ ranges from 7 to 10, depending on the treatment.

Overall, the table shows that the learning rules considered entail different predictions most of the time. This is due to our design involving the 4×4 Anti-Coordination game and should give good chances to estimations of learning types based on decisions alone. We will see below that, despite this fact, the estimates are still significantly biased if only action choices are considered.

	Treatments					
	N-1			F-1		
	RL	PBI	MBR	RL	PBI	MBR
PBI	11			11		
MBR	9	14		10	13	
FL	7	11	9	8	11	10

	Treatments					
	N-2			F-2		
	RL	PBI	MBR	RL	PBI	MBR
PBI	11			15		
MBR	9	14		8	16	
FL	8	13	8	7	14	7

	Treatments					
	N-3			F-3		
	RL	PBI	MBR	RL	PBI	MBR
PBI	11			11		
MBR	11	15		9	15	
FL	8	11	11	8	12	10

Note: Each cell contains the number of rounds two types predict different action choices (out of 19).

Table 4: Separation between learning types on basis of decisions.

4.2 Identifying Learning Rules from Information Requests

Apart from choices we also observe participants' information requests. Naturally, different learning rules imply different needs in terms of information. First, a reinforcement learner only needs to consult her own past payoffs. Else, reinforcement learning is not possible. Similarly, payoff-based imitation requires participants to consult their first-order neighbors' identities, actions, and payoffs, while a myopic best responder has to consult her first-order neighbors identities and action choices. Last, forward-looking learners should look up her first- and second-order neighbors' identities and action choices. Table 5 summarizes this information.

Now if participants consult all the information all the time, observing information searches would not allow to discriminate among the learning rules from these searches. This problem is mitigated in our design, since subjects had to pay for each piece of information they asked for. Table 14 in

		Learning Type			
Info	Neighbor	*RL*	*PBI*	*MBR*	*FL*
Id	1		x	x	x
	2				x
Action	1		x	x	
	2				x
Payoff	Own	x	x		
	1		x		

Table 5: Minimal Information Required for Each Rule (x indicates that a piece of information is required for the corresponding learning rule).

Appendix shows that the monetary cost of information plays an important role in our experimental design. Without such costs, information search would have no discriminatory power.

On the other hand, one may argue that the searches may affect the way people learn and play, but the comparison between the costless and costly endogenous information treatments (as well as the comparison between N and F treatments) reveals that subjects' behavior in unaffected by whether the information is costless or costly (endogenous or full). Imposing a small cost on information, though, makes people focus on the information that is relevant to their decision making process.

Another question is whether participants can trade-off different pieces of information. One could imagine, for example, that a participant asks for choices of her first- and second-order neighbors and then uses this information together with the payoff matrix to compute the payoffs of her first-order neighbors. Clearly, we cannot avoid this, but we believe that (i) making such inferences about the network structure is virtually impossible and (ii) setting such a low cost of consulting actions and payoffs (1 ECU) makes the inferences about not requested information more costly than asking for the desired information directly.

4.3 Decisions and Information Requests in the Data

In this subsection, we provide a descriptive overview of the performance of the different learning rules in the data. Figure 5 shows the fractions of individuals who *simultaneously* request the minimal necessary information corresponding to a rule and choose as prescribed by that rule. The figure illustrates the fractions of subjects, who look up the minimal information set and choose according to each learning rule at least 25% (50 %) of periods, i.e. more than 5 (10) times (left and right panel, respectively). The general insight from these figures is that behavior corresponding to payoff-based imitation is virtually non-existent in the experiment and only few individuals seem to behave as forward-looking agents, while most individuals' searches are consistent with reinforcement learning or myopic best responses.

Figure 6 provides a different look at the data. For each individual and each rule, we calculate the number of times the individual behaves according to this rule (in terms of both information searches and actions taken). For each individual, we then select the rule which is the most frequent predictor of her behavior. In case of ties, we classify the individual as each of these types with uniform shares.[7]

[7] One, two and three individuals have never satisfied the condition for any rule in $N-1$, $N-2$ and $N-3$, respectively.

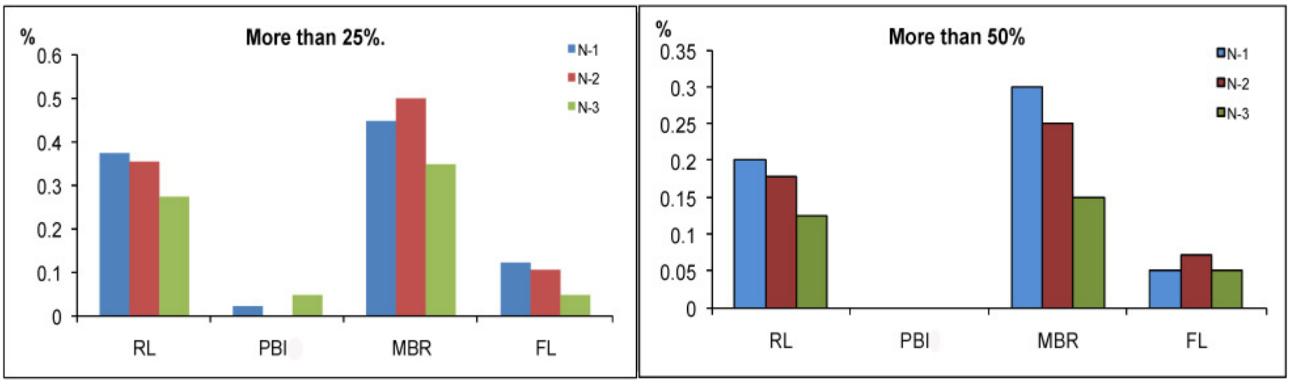


Figure 5: Fractions of subjects, who request the minimal information set *and* play the action as prescribed by each learning type more than 25% (left) or 50% (right) of rounds.

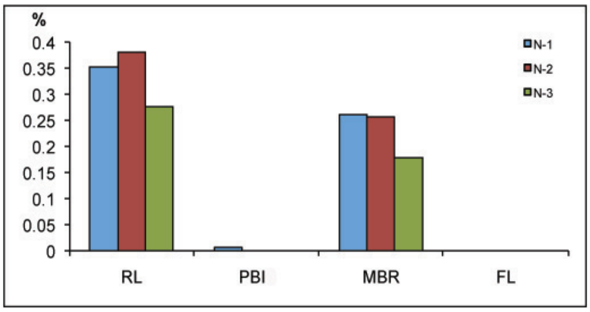


Figure 6: Frequencies of learning types using a simple classification procedure.

This way, each subject is classified into one rule (or with uniform probabilities into various rules) and Figure 6 plots the shares of learning types in the population. We observe that the view is even starker than in Figure 5: virtually no subject in the whole sample is classified as PBI or FL. On the other hand, the two remaining rules seem to play an important role in the data.

Note that Figures 5 and 6 are descriptive and provide only a coarse view of the data. In Figure 6, for example, participants may be categorized as being of a certain type even though they behave in accordance with any rule in only very few periods. In the following section, we estimate learning types more rigorously using maximum likelihood methods.

5 Maximum-Likelihood Estimations

In this section, we introduce the econometric framework and report our main results.

5.1 Econometric Framework

We start specifying our econometric approach. First, the following definition links information searches to learning rules:

Occurrence: In every round a participant requests at least the minimal information she needs to identify the action choice corresponding to her learning type.

These individuals are not included in Figure 6.

While this assumption seems quite innocuous, it can still be too strict in some cases and we will relax it sometimes. For instance, after convergence has occurred participants may not always ask for the minimal information.

CCB remark that Occurrence could be satisfied by chance in their experiment, resulting in low discriminatory power. As discussed above this problem is mitigated in our design, since subjects had to pay for each piece of information they asked for (see Section 4.2 and Table 14 in Appendix).

For each subject i and learning type $k \in \{1, 2, ..., K\}$, we compute the percentage of times subject i asked for the minimum information required for learning rule k ("Compliance with Occurrence"). We sort the resulting percentages into three categories: (i) Z, 0 compliance with Occurrence; (ii) M, $1\% - 49\%$ compliance with Occurrence; (iii) H, $50\% - 100\%$ compliance with Occurrence. This categorization has two advantages. It minimizes the need for structural restrictions (see CCB for more details), while it allows us to evaluate whether subjects who frequently ask for the minimal information set are more precise in their decisions.

In each round a subject's learning type determines her information search (possibly with error) and her type and information search then determine her decision (again possibly with error). Let θ_{kj} denote the probability that a subject has compliance j with rule k in the experiment, where $j \in \{Z, M, H\}$ and $\sum_j \theta_{kj} = 1$ for each k. Note that for a given subject in a given round, a learning type may predict more than one possible action. We assume that in this case participants choose uniformly at random among those actions. Let c denote the number of possible action choices predicted by a learning rule in a given round, with $c \in \{1, 2, 3, 4\}$. A subject employing rule k normally makes decisions consistent with rule k, but in each round, given compliance j she makes an error with probability $\varepsilon_{kj} \in [0, 1]$. We assume that the error rates are independent and identically distributed across rounds and subjects. In the event of an error we assume that participants play each of the four actions with probability $\frac{1}{4}$. As a result, given j and c the probability of making one of the c actions consistent with rule k (either by mistake or as a result of employing rule k) is

$$(1 - \varepsilon_{kj})\frac{1}{c} + \frac{\varepsilon_{kj}}{4} = \left(1 - \frac{4-c}{4}\varepsilon_{kj}\right)\frac{1}{c}, \tag{1}$$

and the probability to choose one of the actions that is *not* consistent with rule k, $\frac{\varepsilon_{kj}}{4}$. For each learning rule k in each period we observe whether or not a participant took a decision consistent with k.

Let $\theta_k = (\theta_{kZ}, \theta_{kM}, \theta_{kH})$ and $\varepsilon_k = (\varepsilon_{kZ}, \varepsilon_{kM}, \varepsilon_{k,H})$, respectively, be the vectors of compliance levels and error rates for each $k \in \{1, 2, ..., K\}$. Let T^{ic}_{kj} denote the number of rounds in which subject i has c possible action choices consistent with rule k and compliance j with learning type k. x^{ic}_{kj} denotes the number of rounds in which i has c possible action choices according to type k, compliance j with k and takes one of the decisions consistent with k. Define $\sum_c T^{i,c}_{kj} = T^i_{k,j}$ and $\sum_c x^{i,c}_{kj} = x^i_{kj}$ for all i, k and j; $x^i_k = (x^i_{kZ}, x^i_{kM}, x^i_{kH})$, $T^i_k = (T^i_{kZ}, T^i_{kM}, T^i_{kH})$; $T^i = (T^i_1, ..., T^i_K)$, $x^i = (x^i_1, ..., x^i_K)$; and $\Im = (T^1, ..., T^N)$ and $X = (x^1, ..., x^N)$. As a result, the probability of observing sample x^i_k and T^i_k is

$$L^i_k(\varepsilon_k, \theta_k | T^i_k, x^i_k) = \prod_j \prod_c \theta_{kj}^{T^{i,c}_{kj}} \left[\left(1 - \frac{4-c}{4}\varepsilon_{kj}\right)\frac{1}{c}\right]^{x^{ic}_{kj}} \left(\frac{\varepsilon_{kj}}{4}\right)^{T^{ic}_{kj} - x^{ic}_{kj}} \tag{2}$$

and the log-likelihood function for the entire sample is

$$\ln LF\left(p, \varepsilon, \theta | \Im, X\right) = \sum_{i=1}^{N} \ln \left\{ \sum_{k=1}^{K} p_k \prod_j \prod_c \theta_{kj}^{T^{ic}_{kj}} \left[\left(1 - \frac{4-c}{4}\varepsilon_{kj}\right)\frac{1}{c}\right]^{x^{ic}_{kj}} \left(\frac{\varepsilon_{kj}}{4}\right)^{T^{ic}_{kj} - x^{ic}_{kj}} \right\}. \tag{3}$$

We assume that our data set is a sample generated by (3).[8] Our aim is to find a mixture model - $p = (p_1, p_2, ..., p_K)$ - that provides the best evidence in favor of our data set. It has been shown that the maximum likelihood method, under mild conditions satisfied by (3), produces consistent estimators for finite mixture models (Leroux, 1992). With K learning types, we have $(6K - 1)$ free independent parameters: $(K - 1)$ independent probabilities p_k, $2K$ type k compliance j information search probabilities θ_{kj}, and $3K$ type-dependent compliance level error rates ε_{kj}. It is well known that it is easy to overparameterize this sort of finite mixture models. Standard information criteria for model selection, such as the Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC), might not perform satisfactorily (Prasad et al., 2007, Biernacki et al. 2000; Cameron and Trivedi, 2010). In the following, we describe how we proceed with model selection (i.e. selection of components) in our case.

For given k, j and c, x_{kj}^{ic} exerts a significant positive influence on the estimated value of p_k as long as the following inequality holds:

$$\ln\left[\frac{(1 - \frac{4-c}{4}\varepsilon_{kj})\frac{1}{c}}{\frac{\varepsilon_{kj}}{4}}\right] \geq 0. \tag{4}$$

The left hand side of (4) is decreasing in the error rate, approaching 0 as $\varepsilon_{k,j}$ tends to 1. This means that type k decisions are taken as evidence of learning rule k only if the estimated error rates suggest that the decisions were made on purpose rather than by error. CCB show that, regardless of the level of compliance j, the log-likelihood function favors type k when T_{kj}^i and hence the estimated θ_{kj} are more concentrated on particular level of compliance j. CCB use the *unrestricted* estimates of θ_{kj} as a diagnostic, giving more confidence to the estimated values of p_k for which both T_{kj}^i and θ_{kj} are *"more"* concentrated on *"high"* levels of compliance j. If, however, the decisions consistent with type k occur with the wrong information search, they are not taken as evidence of type k. We follow their criteria.

Let us comment somewhat on the interpretation of $\theta_k = (\theta_{kZ}, \theta_{kM}, \theta_{kH})$. A high concentration at zero compliance may lead to a probability θ_{kZ} very close to 1, and to a high estimated frequency p_k. However, such high estimated value of θ_{kZ} and, consequently, low estimated values of θ_{kM} and θ_{kH} indicate that subjects do not consult the minimum information corresponding to rule k very often. As a result, it would be hard to argue that learning rule k explains the behavior of the subjects clustered in component k.

With these considerations in mind, we will use the estimated values of θ_k and the error rates ε_k as a tool for selecting the components of our finite mixture model. We will proceed as follows:

Step (a) Estimate the model using K learning rules.

Step (b) If there is a learning type l whose estimated θ_{lZ} is larger than 0.60, go to step (c). If every learning type has an estimated θ_{kZ} smaller than 0.60 (i.e. that the minimal information set was requested *at least* with probability 0.4) and if estimated error rates increase as compliance decreases, stop the estimation process.

Step (c) Remove learning type l, set $K = K - 1$ and go back to step (a).

[8] Our model is an incomplete data structure model. The missing data constitute an unobservable matrix Z whose components are K-dimensional vectors, such that $z_{ik} = 1$ if and only if the underlying generating process of data x^i is a learning type k process.

	Learning types			
Parameters	*RL*	*PBI*	*MBR*	*FL*
K=4				
p_k	0.21	0.62	0.17	0
$\theta_{k,Z}$	0.03	**0.99**	0.05	-
K=3				
p_k	0.20		0.23	0.57
$\theta_{k,Z}$	0.09		0.07	**099**
K=2				
p_k	0.57		0.43	
$\theta_{k,Z}$	0.55		0.10	

Table 6: Selection Algorithm. Treatment N-1

5.2 Estimation Results with Information Searches

We start by illustrating how our algorithm selects learning rules in our treatments with endogenous information. Table 6 shows the results for treatment $N-1$. For the sake of brevity we moved the tables corresponding to treatments $N-2$ and $N-3$ to Appendix.

Table 6 shows the estimated type frequencies p_k and parameters θ_{kZ}. After step (a) $\widehat{\theta}_{PBI,Z} = 0.99$ (in bold in Table 6), meaning that subjects classified as PBI almost certainly do *not* consult the information required by this learning rule. Therefore, our selection criterion suggests that there is no evidence that subjects' behavior was induced by the PBI learning rule, this is also consistent with the evidence in Section 4, and we remove PBI from the estimation. In this way, we iterate steps (a), (b) and (c) as long as there exist no k such that $\widehat{\theta}_{kZ} > 0.6$. The algorithm stops with only two rules, RL and MBR, remaining. Our selection algorithm converges to the same learning rules in $N-2$ and in treatment $N-3$ it selects one learning type: MBR (see Tables 18-20 in Appendix). We describe the result in more detail below.

There is additional information that can be gained by studying Tables 6 and 18-20. In $N-3$, for example, our population is overall best described by MBR. But small percentages of decisions are also very accurately described by other rules that eventually get eliminated by the algorithm. For example 3% are very accurately described by forward-looking with $\theta_{FL,Z} = 0.49$. It is also noteworthy that about 10% of decisions are *accurately* described by RL (step $K = 3$, where $\theta_{RL,Z} = 0.16$). Hence, while (using our selection algorithm) we force the estimation to explain *all* decisions (by the entire population) attributing a significant share of decisions to noise or errors, studying the sequence of estimations can also give us insights into which rules are able to explain accurately and which rules can best account for the more noisy decisions.

We also address carefully the robustness of our results. First of all, we varied the threshold for θ_{kZ}. If we set that threshold to 0.5, we select the same models in treatments $N-2$ and $N-3$. In $N-1$, the algorithm stops with one unique learning rule, MBR. Note, however, that $\widehat{\theta}_{RL,Z} = 0.55$ in Table 6, not that far from 0.5. If we increase the threshold up to 0.9, the algorithm would select the same population composition in the three treatments, being the percentages of RL and MBR very similar in the three networks (see Table 7).[9]

[9] This finding may reinforce our general conclusion that learning types composition is relatively stable across different networks. Nevertheless, we still use a more conservative threshold $\widehat{\theta}_{k,Z} \leq 0.6$ as our benchmark, because we believe

	Treatments					
	N-1 ($LL = -1406$)		N-2 ($LL = -2019$)		N-3 ($LL = -1474$)	
Parameters	RL	MBR	RL	MBR	RL	MBR
p_k	0.57	0.43	0.59	0.41	0.68	0.32
θ_Z	0.55	0.10	0.48	0.14	**0.90**	0.65
θ_M	0.06	0	0.09	0	0.10	0.11
θ_H	0. 39	0.90	0.43	0.86	0	0.23
ε_Z	1	1	1	1	1	1
ε_M	0.52	-	0.5	-	0.58	0.74
ε_H	0.51	0.46	0.55	0.41	-	0.57

Table 7: Information Search and Decisions

Second, we artificially altered the order of elimination of the learning types (for which the minimal information set was rarely requested) and in all cases we converge to the same mixture composition as in the benchmark case. Hence, all our results are robust to the order of elimination of learning types.[10]

Next we describe our main results. Table 7 reports the maximum likelihood estimates of learning type probabilities, p_k, unconditional compliance probabilities, θ_{kj}, and compliance conditional error rates, ε_{kj}, in the selected models. Figure 7 illustrates the estimated frequencies for reinforcement learning and myopic best-response learning.

In treatment $N-1$, 57% of the population are best described as reinforcement learners and the remaining 43% as myopic best responders. RL has high compliance with occurrence ($\hat{\theta}_{RL,H} = 0.39$), while $\hat{\theta}_{MBR,H}$ even equals 90%. In both cases, estimated error rates increase as compliance decreases (i.e. the more frequently people classified into each rule consult the information the more they act in harmony with the rule) and the estimated unconditional *zero* compliance probabilities are lower than the corresponding error rates. These results suggest that the estimated type frequencies of RL and MBR are highly reliable.

In $N-2$, 59% and 41% of subjects are best described by RL and MBR, respectively. The estimated θ's and ε's are also well behaved. Note that the estimates are remarkably similar in $N-1$ and $N-2$. In both networks, the combination of reinforcement learners and myopic best responders best describes the population, being the estimates for myopic best responses more accurate and reinforcement learning able to absorb somewhat more of "noisy behavior".

Finally, for the stricter threshold of 60% only MBR survives in $N-3$. At step $K = 2$, 68% of subjects are classified as reinforcement learners, but they consult their corresponding information set with probability 0.1. Consequently, we remove RL and end up with MBR only.

In sum, the topology of the underlying network seems to have a limited influence on subjects' learning types. The mixture composition is very similar in $N-1$ and $N-2$, with a majority

that people classified into one learning type should request the information corresponding to this rule more often than in 10% of cases.

[10]We have also considered some variants of the learning rules to see how robust our results are to their definitions. See Section 6 for the results.

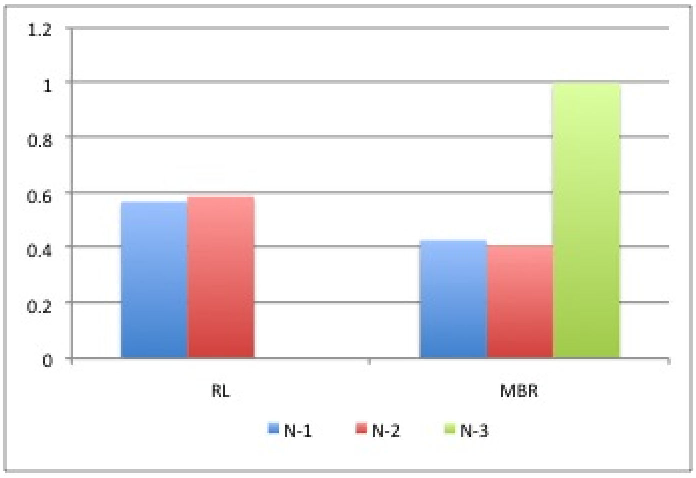


Figure 7: Estimation of p_k using information requests and decisions.

of subjects best described as reinforcement learners and the remaining participants as belief-based learners. MBR seems to play a more important role in $N-3$. Here, the whole population is classified into belief-based learning types and there is no evidence for reinforcement learning. This is possibly due to the fact that learning in $N-3$ is less complex cognitively, because *(i)* there are no cycles and *(ii)* many players only have one neighbor in this network. As a consequence, players may resort to more sophisticated rules (such as MBR) more often compared to a simpler rules such as RL. We will come back to this issue below. However, we have to be cautious about these results, because standard errors are not available. This makes it impossible to test network effects formally.

Since almost all participants can be described by either reinforcement learning rules or belief based rules, our results support the assumptions of EWA (Camerer and Ho, 1998; Camerer et al., 2002), which includes reinforcement and belief-based learning as special cases as well as some hybrid versions of the two. Unlike in EWA we do not restrict to those models ex ante, but our results suggest that - at least in the context considered - a researcher may not be missing out on too much by focusing on those models. While EWA should be a good description of behavior at the aggregate level, at the individual level only about 16% of our participants request information consistent with *both* reinforcement learning and belief-based learning rules.

There might be two caveats regarding our results. First, could it be that we are overestimating the frequency of RL, because participants might look up their own payoffs just because they want to know their payoffs and not because they use this information in their learning rule? We probably do, but only to a small extent. Note, first, that the estimation procedure identifies high correlations between information searches and "correct" choices given the learning models consistent with the information search. As a result, if a decision-maker always looks up some information for other reasons (unrelated to the way she learns and play), then this will *not* lead to high correlations and hence will *not* mislead the estimation procedure. In addition, the fact that we do not find evidence for RL in $N-3$ indicates that this is a minor issue in our study.

Another caveat could be that we are not giving imitation learning the best chances here, since players are not symmetric in terms of their network position and since players typically want to choose a different action than their neighbors in an Anti-Coordination game. To address this issue, we verify whether imitation of second-order neighbors (whose actions players typically want to mimic in equilibrium) can best describe the behavior of any subject. We allow for both conformist and payoff-based imitation and observe that no subject is best described solely by imitating within the second-order neighborhood.[11]

A Coordination game would certainly have given better chances to imitation learning. However in these games our learning models are most often indistinguishable in terms of the choices they imply for any given information search. Since our main focus is the comparison of estimations with and without taking into account information searches, we designed the experiment in such a way to give the latter the best possible chances. If one was primarily interested in understanding in which situations agents resort to social learning (imitation) as opposed to best response or reinforcement learning, then one would need to conduct additional experiments involving different games. We leave these issues for future research.

5.3 Estimation Results without Information Searches

How important role does monitoring of information requests play in our estimations? To address this question, in this section we estimate learning types solely on basis of participants' observed choices (*disregarding* their information searches). The objective is to show that if we only use information about subjects' behavior the estimates are less accurate, in spite of the fact that our design should give these estimations good chances (see Section 4.1).

Recall that we assume that a type-k subject normally makes a decision consistent with type k, but she can make an error with probability ε_k, in which case she chooses an action with probability $\frac{1}{4}$. Let x_k^{ic} measure the number of rounds in which subject i has c possible action choices and takes a decision consistent with k. Under this model specification the probability of observing sample x_k^i can be written as

$$L_k^i\left(\varepsilon_k|x_k^i\right) = \prod_{c=1,2,3,4}\left[\left(1-\frac{4-c}{4}\varepsilon_k\right)\frac{1}{c}\right]^{x_k^{i,c}}\left(\frac{\varepsilon_k}{4}\right)^{T_k^{i,c}-x_k^{i,c}}. \tag{5}$$

Adding up over all individuals and learning types, we get the log-likelihood of the whole data set:

$$\ln LF\left(p,\varepsilon|x\right) = \sum_{i=1}^{N}\ln\left\{\sum_{k=1}^{K}p_k\prod_{c=1,2,3,4}\left[(1-\frac{4-c}{4}\varepsilon_k)\frac{1}{c}\right]^{x_k^{i,c}}\left(\frac{\varepsilon}{4_k}\right)^{T_k^{i,c}-x_k^{i,c}}\right\}. \tag{6}$$

As in (3), the influence of x_k^{ic} on the estimated value of p_k decreases as ε_k tends to 1 and meaning that learning type-k decisions are thus taken as evidence of rule k only to the extent that the estimated value of ε_k suggests they were made on purpose rather than in error.

Parameters of model (6) are estimated using maximum likelihood methods as before. In this case we have $2K-1$ free independent parameters, $(K-1)$ corresponding to frequency types p_k, and K corresponding to the error rates. Since under this specification the shape of the objective function (6) is better behaved compared to (3), we can now estimate the standard error rates. Table 8 reports estimated frequencies and error rates for each treatment.

Note that in all cases we have evidence in favor of the four learning rules. Based on these results we could conclude that there is evidence of payoff-based imitation (8% in $N-1$, 4% in $N-2$ and 23% in $N-3$) and forward-looking learning (30% for $N-1$, 11% and 5% for $N-2$ and $N-3$ respectively). However, our data in Section 4 show that subjects hardly ever checked the necessary

[11]Two subjects are simultaneously best described by more than one learning rule in $N-3$, including conformist imitation within the second-order neighborhood, reinforcement learning and myopic best responding. However, one of them is best described by each of these rules in only one period, while the other subject is comply with them only three times. These figures suggest that their compliance with imitation (as much as with the other rules) is rather accidental than made on purpose.

information to identify the corresponding action choices. Consequently, it is very unlikely that these learning rules have generated the behavior of subjects in the experiment. Summing across those learning rules, these numbers indicate that 15% of participants are mis-classified if we only consider action choices and abstract from which information they consult!

The important message of this section is that if we disregard the information that subjects request we may end up accepting learning rules that subjects actually do not use. Remember also that our design (involving the 4×4 Anti-Coordination game) was chosen in order to give estimation by choices alone good chances to detect learning behavior, since learning rules can be discriminated better by focusing on choices alone. Hence in general the biases resulting from estimations by decision alone may be much more severe for smaller and coordination-like games than those that we encounter.

Treatment $N-1$ ($LL=-760$)				
	RL	$PBI1$	MBR	FL
p_k	0.21***	0.08*	0.42***	0.30
(st.er.)	0.07	0.06	0.10	
ε_k	0.08***	1***	0.58***	0.42***
(st.er.)	0.03	0.10	0.05	0.054
Treatment $N-2$ ($LL=-1022$)				
	RL	$PBI1$	MBR	FL
p_k	0.49***	0.04***	0.35***	0.11
(st.er.)	0.01	0.08	0.01	
ε_k	0.26***	0.48***	0.47***	0.76***
(st.er.)	0.003	0.02	0.01	0.05
Treatment $N-3$ ($LL=-772$)				
	RL	$PBI1$	MBR	FL
p_k	0.51***	0.23***	0.21***	0.05
(st.er.)	0.01	0.01	0.01	
ε_k	0.34***	0.70***	0.42***	0.28***
(st.er.)	0.01	0.01	0.01	0.01

Note: (***) Significant at 1% level; (**) at 5% level; (*) at 10% level.

Table 8: Estimation based solely on observed behavior.

How do we know that the model with information search gives “better” and not just “different” estimates than the model without information search? Obviously estimations that take into account information searches use more information and hence they can rule out learning rules that are plausible when looking at decisions only, but simply not possible because the decision-maker did not have the minimal information needed for that rule. The estimation procedure identifies high correlations between information search and “correct” choices given the learning models consistent with the information search. Hence if a decision-maker always looks up some information for other reasons (unrelated to the way she learns), then this will *not* lead to high correlations and hence will *not* mislead the procedure based on information searches. The only case in which the process with information search could be misled is if (i) two different rules predict the same choices and (ii) information needed for one rule can be deduced from information needed for the other rule. Our experimental design renders (ii) unlikely, and Table 4 shows that (i) is only very rarely the case in our experiment. Note also that situations such as (i) will likely affect estimations that disregard information searches even more strongly.

6 Further Results

In this section, we estimate some alternative models. First, we estimate rules by player position to see whether network position (central vs. peripheral) affects how agents learn. Second, we substitute the MBR rule with different variations of belief-based learning. Third, we provide additional estimations making stronger distributional assumptions. Fourth, using simulated data we evaluate to what extent our econometric model is capable of identifying the learning rules present in the population. Finally, we relax our assumption of occurrence.

6.1 Estimation by Player Position

We estimate our model separately for different player positions in the network to understand whether how people update their behavior is influenced by their position in the network. To estimate the model separately for each position in the networks would lead to very small samples (of 5-8 independent observations only) and hence very likely to small sample biases. To mitigate this problem we categorize players into two different categories according to whether they have a "central" or "peripheral" position in the network. In Network 1 all players are symmetric and hence have the same position. As a consequence, we focus on networks 2 and 3 in this analysis.

In network 2 we call Players 1, 2, 7 and 8 "peripheral" and Players 3, 4, 5 and 6 "central" (see Figure 2), while in network 3 we call Players 1, 5, 6, 7 and 8 "peripheral" and Players 2, 3 and 4 "central" (Figure 3).

In $N-2$ we have different results for "peripheral" and "central" players. All peripheral players are classified as MBR learners, while 36% of central players as RL and 64% as MBR. The intuition is similar to before. Peripheral players have only few neighbors (in most cases only one) and hence it is cognitively less demanding for peripheral players to store information about past play of their neighbors. Hence, it seems intuitive that players with fewer neighbors rely more often on myopic best response that is computationally more demanding compared to RL. Player position does also have an effect on *behavior*, as we discuss in more detail in our companion paper (Kovarik et al, 2011). Figure 8 shows these results and compare them with the benchmark estimations for $N-2$.

In $N-3$ players' position has virtually no effect on the learning rules players are best described by. Both "peripheral" and "central" players (as well as the benchmark estimation from previous section) are classified as MBR.

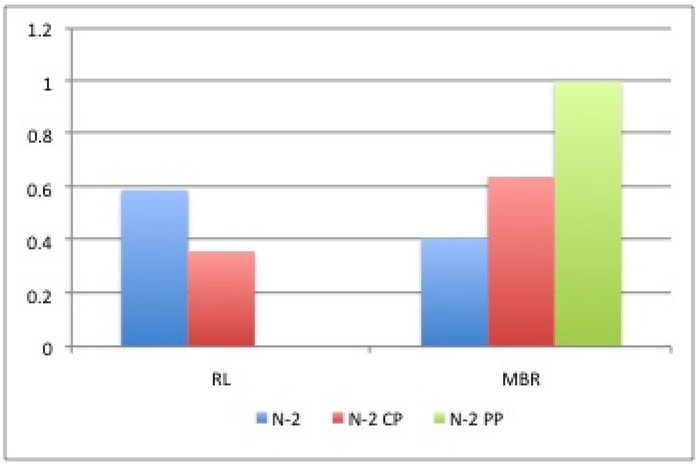


Figure 8: Estimation by Player Position in treatment N-2. $N-2$ CP refers to "central" players 3,4,5 and 6; $N-2$ PP includes "peripheral" players 1,2,7 and 8 (see Figure 2).

6.2 Splitting the Sample by Cognitive Reflection

In this section, we split our subjects according to their answers in a cognitive reflection test (Fredericks, 2005), conducted in a post-experimental questionnaire. The test consists of the three following questions

1. A bat and a ball cost Euro 1.10 in total. The bat costs Euro 1.00 more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

All these questions have an answer that immediately springs into mind (10 cents, 100 minutes, 24 days), but which is wrong. The right answer (5 cents, 5 minutes, 47 days) can only be found by engaging in some cognitive reflection. Note that the test is not measuring intelligence, but rather the willingness of subjects to engage in costly cognitive reflection.

In order to avoid small sample biases (and since the estimates are very similar in $N-1$ and $N-2$), we aggregate our data from treatments $N-1$ and $N-2$ and subsequently split this sample by using the following measure of cognitive reflection. We categorize participants as displaying "low" cognitive reflection if they answered 0 or 1 question correctly and participants as displaying high cognitive reflection if they answered 2 or 3 questions correctly. We then estimate the distribution of learning types as above separately for "low" and "high" cognitive reflection.

	Cognitive Reflection			
	Low ($LL=-1479$)		High ($LL=-1876$)	
Parameters	RL	MBR	RL	MBR
p_k	0.51	0.49	0.57	0.43
θ_Z	0.92	0.72	0.48	0.13
θ_M	0.08	0.05	0.06	0.00
θ_H	0.00	0.22	0.46	0.87
ε_Z	1	1	1	1
ε_M	0.47	0.71	0.59	-
ε_H	-	0.41	0.48	0.39
Observations	798		1026	

Table 9: Information Search and Decisions with the Sample split by Cognitive Reflection.

Table 9 reports the results of this exercise. The results center around the same two learning rules as above, namely reinforcement learning and myopic best response. The estimated distribution of these two types is remarkably similar across our two categories. Interestingly, though, the precision of estimates is much higher for the sub-sample categorized as "high cognitive reflection". Presumably hence participants that are more willing to engage in cognitive reflection display less noisy information searches as well as choices which is picked up by the precision of our estimates.

6.3 Fictitious Play with Limited Recall

In this section we estimate model (3) with different alternative variations of belief learning. In particular, we assume that subjects only base their learning on a limited number of last periods. This way, myopic best responders are one end of this classification basing their decisions on the last round only. We consider six alternative specifications: players best respond to the play of their opponents in the last three, six, nine, twelve, fifteen and twenty past periods to construct their beliefs. Note that the last variation represents the standard fictitious-play learning. Denote by FP_s the variation, under which subject based their decision on the last s periods. Hence, under this terminology, myopic best-responders are fictitious players who only recall the last period, i.e. FP_1, and fictitious players correspond to FP_{20}. We compare these alternatives with the benchmark model and rank them according to their log-likelihood values. For each treatment we solely present results for the best performing model (Table 21 in Appendix).

In all treatments the best-performing model is the benchmark from Section 5.2 with MBR (i.e. FP_1). However, the increment in the log-likelihood value in the benchmark model with respect to the second best-performing model is very small (lower than 1% in all cases).

In $N-1$ there is no difference between the benchmark model and the model with FP_3 and the estimated parameters are remarkably similar. In the other two treatments the model including FP_6 outperforms the other alternative models and the estimated frequency types are again very similar to the benchmark that contain MBR. In all cases FP_{20} is among the last in the ranking.

These results suggest that depending on the environment assuming that FP learners have bounded memory to form their beliefs explains a little more of the variation in subjects' decisions compared to FP_{20} model. Hence, the proposed variants of fictitious play can be empirically more relevant than the standard definition of this learning rule. Note though that we have not considered weighted fictitious play in the model.

6.4 Estimation Using a Poisson Process

In order to asses to what extent our results depend on the distributional assumptions behind the likelihood function, in this section we re-estimate the learning rules assuming that the processes behind our data follow a Poisson distribution.

We keep assuming that a participant should frequently request the minimal information she needs to identify the action choice corresponding to her learning type. Let I_k^i denote the number of rounds in which subject i searches information consistent with learning type k during the experiment and x_k^i denotes the number of rounds in which subject i makes a decision consistent with learning rule k. We assume that the variables I_k^i and x_k^i follow a Poisson distribution with means μ_k and λ_k, respectively. Note that we again assume type-dependent parameters, which takes into account that the difficulty in processing information varies across learning rules.

The probability of observing sample (I_k^i, x_k^i) is

$$L_k^i(\mu_k, \lambda_k | I_k^i, x_k^i) = \frac{e^{-\mu_k} \mu_k^{I_k^i}}{I_k^i!} \frac{e^{-\lambda_k} \lambda_k^{x_k^i}}{x_k^i!},$$

and the log-likelihood function is

$$lnLF(p,\mu,\lambda|I,x)=\sum_{i=1}^{N} ln\left(\sum_{k=1}^{K} p_k \frac{e^{-\mu_k}\mu_k^{I_k^i}}{I_k^i!}\frac{e^{-\lambda_k}\lambda_k^{x_k^i}}{x_k^i!}\right). \tag{7}$$

Because of problems of over-parameterization related to finite mixture models, we apply a selection algorithm similar to that of Section 5. If $\widehat{\mu}_k < 0.5$, then subjects classified as type k hardly ever check the minimal information set consistent with this rule and cannot identify the corresponding action choice. As a result, we remove the types whose $\widehat{\mu}_k < 0.5$. Table 10 shows the estimation results.

Endogenous Information Treatments

	$N-1$ ($LL=-255$)		$N-2$ ($LL=-339$)		$N-3$ ($LL=-114$)	
	RL	MBR	RL	MBR	MBR	FL
p_k	0.58	0.42	0.52	0.48	0.30	0.70
μ_k	1.92	7	2.15	6.34	12.76	0.21
λ_k	1.02	4.41	1.15	3.99	5.68	0

Table 10: Poisson distribution. Estimation based on information search and observed behavior.

Our estimates provide evidence in favor of reinforcement and belief-based learners in $N-1$ and $N-2$. In addition, the estimated frequencies are remarkably similar to model (3) in Table 7. Again, lower values of μ_k and λ_k suggest that RL absorbs more of the noisy behavior.

In $N-3$ we see that 70% of subjects are classified as forward-looking learners if $K=2$. However, the estimated parameter $\widehat{\mu}_{FL}=0.21$; that is, subjects classified as FL hardly ever check the information set corresponding to this rule. Consequently, using the our selection criterion all subjects are classified as MBR learners as in (3).

Hence, this alternative and more restrictive model confirms the type composition of the sample from Section 5. This conclusion still holds if we increase the threshold $\widehat{\mu}_k$ up to (almost) two in the selection algorithm.

6.5 Evaluating the Information Search Model Using Monte Carlo

In this section we analyze the ability of our model specification to detect the true mixture of the population. The exact objective of this exercise is to answer the following question: If the estimated composition of the population in $N-1$ is the true composition, is our estimation procedure able to detect it precisely?

To this aim we simulate a data set with the same structure as our experimental data. We let computers simulate the behavior of two different learning types, RL and MBR. Then, we estimate the model (3) using this data set and apply our selection algorithm to test whether our procedure can recover the true data-generating processes.

In the paper we only report results for treatment $N-1$.[12] To mimic our experiment, we simulate data for five groups of eight players (40 subjects in total) who play our Anti-coordination game for 20 periods. We assume that 58% of subjects are RL and 42% are MBR in all simulations and the types are randomly distributed on the network according to the above probabilities. We consider three different parameter constellations (summarized in Table 11) as follows:

[12]Since the conclusions are the same for $N-2$ and $N-3$, we do not report them here.

1. *Full Compliance* (**FC**): subjects search their respective information set with probability 1 and make no mistake in choosing the corresponding action choice.

2. *High Compliance* (**HC** hereafter): subjects search their corresponding information search with high probability and make mistakes choosing actions with low probability,

3. *Low Compliance* (**LC**): subjects have low compliance with occurrence and make mistakes with high probability.

For each case we have 250 computer-generated samples with these characteristics.

Network 1 (Circle)

Parameters	FC	HC	LC
Numb. RL	23	23	23
Numb. MBR	17	17	17
Total	40	40	40
Types' Parameters ($k = RL, MBR$)			
$\theta_{k,Z}$	0	0	0.55
$\theta_{k,M}$	0	0.15	0
$\theta_{k,H}$	1	0.85	0.45
$\varepsilon_{k,Z}$	1	1	1
$\varepsilon_{k,M}$	0	0	0
$\varepsilon_{k,H}$	0	0.10	0.55

Table 11: Assumptions for Monte Carlo Simulations

Table 12 reports the results of this exercise. We observe that in all cases our selection algorithm identifies correctly the learning rules present in the population. The shares of PBI and FL are virtually zero in all cases. Moreover, we find very small biases in the estimated frequencies in both **FC** and **HC**. Hence, if people are relatively precise both making their choices and looking up the information, our estimation procedure succeeds selecting the population composition in all cases.

As subjects become less precise in their information searches and decisions (**LC**), we still recover which types are present in the population, but there are biases in the estimated values. In our particular case, the mechanism overestimates the presence of MBR by 13% and underestimates the share of RL by the same amount. The intuition of such biases is as follows. First, recall the variable $T^i_{k,Z}$ that measures the number of rounds in which subject i does not consult the information set corresponding to rule k. Notice that in **FC** the variable $T^i_{RL,Z} = T^i_{MBR,Z} = 0$. As subjects become less precise in their information searches, these variables increase, which in turn increases the concentration in zero compliance level. Since in our computer-based samples 23 (out of 40) subjects have $T^i_{MBR,Z} = 19$, this concentration is relatively higher for MBR generating the observed shift toward more MBR at the expense of RL.

Since our estimates in Section 5.2 lie somewhere between the HC and LC scenarios, we select the correct mixture composition of the population and if there exist some biases in the estimates these biases will be relatively small.

Network 1 (Circle)

	RL	PBI	MBR	FL
True p_k	0.575	0	0.425	0
FC				
$\widehat{p}_k$	0.597	0.002	0.399	0.002
(se)	0.07	0.008	0.075	0.006
Bias	0.02	0.002	-0.03	0.0012
HC				
$\widehat{p}_k$	0.5710	0.0004	0.4283	0.0001
(se)	0.0447	0.004	0.043	0.0004
Bias	-0.003	0.004	0.003	0.0001
LC				
$\widehat{p}_k$	0.443	0	0.556	0
(se)	0.0302	0	0.0302	0
Bias	-0.132	0	0.132	0

Table 12: Monte Carlo Simulations

6.6 Relaxing the Assumption of Occurrence

In this section we relax our assumption of occurrence, allowing people not to look up the information every time they play. This is important here, since we have observed that subject consulted the information less frequently after convergence to an equilibrium.

To this aim, we assume that a subject has the information she needs to identify the action choice corresponding to her type if she has asked for the minimal information set *at least once in the last four periods.* Table 13 reports the estimates, which again confirm our results in Section 7. In $N-1$ and $N-2$ we have evidence in favor of RL and MBR and their shares are relatively stable in both cases. In $N-3$ we observe the major difference, we now have evidence in favor of RL and MBR. In all case the estimated θ's and ε's are well behaved, hence we put confidence in the estimated frequency types.

	Treatments					
	N-1 ($LL = -1081$)		N-2 ($LL = -1408$)		N-3 ($LL = -1179$)	
Parameters	RL	MBR	RL	MBR	RL	MBR
p_k	0.48	0.52	0.49	0.51	0.43	0.57
θ_Z	0.14	0.03	0.11	0.02	0.21	0.04
θ_M	0	0	0	0	0	0
θ_H	0.86	0.97	0.89	0.98	0.79	0.96
ε_Z	1	1	1	1	1	1
ε_M	-	-	-	-	-	-
ε_H	0.66	0.41	0.44	0.51	0.41	0.69

Table 13: Information Search and Decisions

7 Concluding Remarks

We study how subjects learn in a 4×4 normal form (anti-) coordination network game. The way in which subjects interact and the way in which information spreads through the network may shape subjects' decisions. To gain insight into these issues we vary the topology of the underlying interaction structure. In particular, we systematically vary the heterogeneity in degree, holding constant other networks characteristics. As a second treatment variation we let subjects decide which information they would like to see before making their decisions or provide them with full information.

We find that there is substantial heterogeneity in the way people learn in our data. However most agents can be classified as either reinforcement learners or belief learners. Our results suggest that in context of complex environments (due to more neighbors or a more complex network structure), subjects rely more often on learning rules (such as reinforcement learning) that are cognitively less demanding compared to belief-based learning models. Future research is needed to address the question of heterogeneity and context stability across different games and other contexts and at the individual rather than aggregate level.

References

[1] Alós-Ferrer, C. and K. Schlag (2009), Imitation and Learning, chapter in Handbook of Rational and Social Choice , P. Anand, P. Pattanaik and C. Puppe (eds.), Oxford University Press.

[2] Alós-Ferrer, C. and S. Weidenholzer (2008), Contagion and Efficiency, Journal of Economic Theory 143, 251-274.

[3] Bendor, J., Mookerherjee, D., and Ray, D. (2001), "Aspiration-Based Reinforcement Learning in Repeated Interaction Games: An Overview," International Game Theory Review 3, 159-174.

[4] Biernacki C., Celeux G., and Govaert G., (2000), "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood," IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 7, 719-725.

[5] Blume, L. (1993), The statistical mechanics of strategic interaction, Games and Economic Behavior, 5, 387-424.

[6] Blume, L. (2004), "Evolutionary Equilibrium with Forward-Looking Players", working paper Santa Fe Institute.

[7] Boyd, R. and P. Richerson (2005), The Origin and Evolution of Cultures, Oxford University Press.

[8] Börgers, T. and R. Sarin (1997), Learning through Reinforcement and Replicator Dynamics, Journal of Economic Theory 77, 1-14.

[9] Boylan, R. and El-Gamal, M. (1993), "Fictitious Play: A Statistical Study of Multiple Economic Experiments," Games and Economic behavior 5, 205-222.

[10] Cabrales, A. and García-Fontes, W. (2000), Estimating learning models from experimental data, Universitat Pompeu Fabra Economics and Business Working Paper No. 501. Available at SSRN: http://ssrn.com/abstract=246526 or doi:10.2139/ssrn.246526.

[11] Camerer, C. (2003), Behavioural Game Theory: Experiments in Strategic Interaction, Princeton University Press.

[12] Camerer, C. and Ho, T.-H. (1998), Experienced Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation, Journal of Mathematical Psychology 42, 305-326.

[13] Camerer, C. and Ho, T.-H. (1999), Experienced Weighted Attraction Learning in Normal-From Games, Econometrica, 67, 827-874.

[14] Camerer, C., Ho, T.-H., and Chong, J.K. (2002), Sophisticated Experience-Weighted Attraction Learning and Strategic Teaching in Repeated Games, Journal of Economic Theory 104, 137-188.

[15] Camerer, C., Johnson, E., Rymon, T., and Sen, S. (1993), "Cognition and Framing in Sequential Bargaining for Gains and Losses," in Binmore, K., Kirman, A. and Tani, P. (eds), Frontiers of Game Theory. Cambridge: MIT Press, 27-47.

[16] Cameron, A. C., and Trivedi, P. K, (2010), "Microeconometrics Using Stata," Stata Press, Ch. 17 p. 599.

[17] Cheung, Y. and Friedman, D. (1997), Individual Learning in Games: Some Laboratory Results, Games and Economic behavior, 19, 46-76.

[18] Corten, R. and K. Cook (2008), Cooperation and Reputation in Dynamic Social Networks, In Poalucci,M. (ed.): Proceedings of the First International Conference on Reputation: Theory and Technology – ICORE 09, Gargonza, Italy.

[19] Costa-Gomes, M., Crawford, V., and Broseta, B., (2001), Cognition and Behavior in Normal-Form Games: An Experimental Study, Econometrica, 69(5), 1193-1235.

[20] Ellison, G. and D. Fudenberg (1993), Rules of Thumb for Social Learning, Journal of Political Economy 101, 612-643.

[21] Engle-Warnick, J. and Ruffle, B. (2005), The Strategies Behind Their Actions: A Method to Infer Repeated-Game Strategies and an Application to Buyer behavior, mimeo, McGill University.

[22] Erev, I. and Roth, A. (1998), "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," American Economic Review 88, 848-881.

[23] Eshel, I., L. Samuelson and A. Shaked (1998), Altruists, Egoists and Hooligans in a Local Interaction Model, American Economic Review 88, 157-179.

[24] Feltovich, N. (2000), Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games, Econometrica 68(3), 605-641.

[25] Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2): 171-178.

[26] Fudenberg, D. and D.K. Levine (1998), The Theory of Learning in Games, Cambridge: MIT-Press.

[27] Fujiwara-Greve, T. and C.Krabbe-Nielsen (1999), "Learning to Coordinate by Forward Looking Players", Rivista Internazionale di Scienze Sociali CXIII(3), 413-437.

[28] El-Gamal M.A. and Grether, D.M. (1995), Are People Bayesian? Uncovering behavioural Strategies, Journal of the American Statistical Association 90(432), 1137-1145.

[29] Galeotti, A., Goyal, S., Jackson, M., Vega-Redondo, F., and Yariv, L. (2010), Network games, Review of Economic Studies, forthcoming.

[30] Gill, J., and King, G. (2004), What to Do When Your Hessian Is Not Invertible, Sociological Methods and Research 33 (1): 54-87

[31] Goyal, S. (2003), Learning in networks: a survey, Group formation in economics: networks, clubs, and coalitions, Ed. Demange, G. and Wooders, M. Cambridge University Press.

[32] Goyal, S. and F. Vega-Redondo (2005), Network formation and social coordination, Games and Economic Behavior 50, 178-207.

[33] Harless, D.W. and Camerer, C. (1994), The Predictive Utility of Generalized Expected Utility Theories, Econometrica 62(6), 1251-1289.

[34] Ho,T.-H., Wang, X., Camerer, C. (2008), Individual Differences in EWA Learning with Partial Payoff Information , The Economic Journal 118, 37-59.

[35] Hojman, D. and A. Szeidl (2006), Endogenous Networks, Social Games and Evolution, Games and Economic Behavior 55 (1), 112-130.

[36] Hopkins, E. (2002), Two Competing Models of How People Learn in Games, Econometrica 70(6), 2141-2166.

[37] Houser, D., Keane M., and McCabe K. (2004), Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm, Econometrica 72, 781-822.

[38] Jackson, M. O. and A. Watts (2002), On the formation of interaction networks in social coordination games, Games and Economic Behavior 41, 265-291.

[39] Johnson, E., Camerer, C., Sen, S. and Rymon, T. (2002), “Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining,” Journal of Economic Theory, 104, 16-47.

[40] Kandori, M., G. Mailath and S. Rob (1993), ”Learning, mutation, and long run equilibria in games,” Econometrica 61, 29-56.

[41] Kirchkamp, O. and Nagel, R. (2007), “Naive learning and cooperation in network experiments”, Games and Economic behavior 58, 269-292.

[42] Kovarik, J., F. Mengel and J.G. Romero (2011), (Anti-)Coordination in Networks, Meteor Research Memorandum.

[43] Leroux B. G., (1992), “Consistent Estimation of a Mixing Distribution,” The Annals of Statistics 20, 3, 1350-1360.

[44] McCabe K., Mukherji, A. and, Runkle, D. (1993), Learning and Limited Information in Games: An Experimental Study, Federal Reserve Bank of Minneapolis.

[45] McKelvey, R. and Palfrey, T. (1995), ”Quantal Response Equilibria for Normal Form Games”, Games and Economic behavior 10: 6–38.

[46] Mengel, F. (2009), Conformism and Cooperation in a Local Interaction Model, Journal of Evolutionary Economics 19 (3), 397-415.

[47] Mookherjee, D. and Sopher, B. (1994), Learning Behavior in an Experimental Matching Pennies Game, Games and Economic behavior 7, 62-91.

[48] Mookherjee, D. and Sopher, B. (1997), Learning and Decision Costs in Experimental Constant Sum Games, Games and Economic behavior 19, 97-32.

[49] Pingle, M. and Day, R. (1996), Modes of economizing behavior: Experimental evidence, Journal of Economic Behavior and Organization.

[50] Prasad A. N., Shi P., and Tsai Chih-Ling, (2007), “Extending the Akaike Information Criterion to Mixture Regression Models,” Journal of the American Statistical Association 102, 477, 244-254.

[51] Roth, A.E. and I. Erev (1995), Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term, Games and Economic Behavior 8, 164-212.

[52] Schlag, K. (1998), Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-Armed Bandits, Journal of Economic Theory 78, 130-56.

[53] Schoenberg, R. (2001), Constrained Optimization, Aptech Systems, Inc. Maple Valley, WA.

[54] Shachat, J. and Walker, M. (2004), ”Unobserved Heterogeneity and Equilibrium: An Experimental Study of Bayesian and Adaptive Learning in Normal Form Games,” Journal of Economic Theory 114, 280-309.

[55] Skyrms, B. and R. Pemantle (2000), A dynamic model of social network formation, PNAS 97, 9340-9346.

[56] Vega-Redondo, F. (2003), Economics and the Theory of Games, Cambridge University Press.

[57] Wilcox, N., (2005), Theories of Learning in Games and Heterogeneity Bias: Some Nonsolutions, Unpublished Manuscript, University of Houston.

[58] Wilcox, N., (2006), Theories of Learning in Games and Heterogeneity Bias, Econometrica, 74(5), 1271-1292.

[59] Young, P. (1998), Individual Strategy and Social Structure, Princeton: Princeton University Press.

A Appendix: Nash Equilibria of Network Game

Nash equilibria		
Network 1	Network 2	Network 3
(A,B,A,B,A,B,A,B)	**(A,B,A,B,B,A,A,A)**	**(A,B,A,B,A,A,A,A)**
(B,A,B,A,B,A,B,A)	**(B,A,B,A,A,B,B,B)**	**(B,A,B,A,B,B,B,B)**
(C,D,C,D,C,D,C,D)	**(C,D,C,D,D,C,C,C)**	**(C,D,C,D,C,C,C,C)**
(D,C,D,C,D,C,D,C)	**(D,C,D,C,C,D,D,D)**	**(D,C,D,C,D,D,D,D)**
(D,D,C,D,C,D,D,C)	(C,D,D,D,C,D,C,C)	(D,C,D,D,C,C,C,C)
(D,C,D,C,D,D,C,D)	(D,C,D,D,C,D,C,C)	(C,D,D,C,D,D,D,D)
(C,D,C,D,D,C,D,D)	(D,C,D,D,D,C,C,C)	(A,B,C,A,B,B,B,B)
(D,C,D,D,C,D,D,C)	(C,D,D,C,C,D,D,D)	(D,C,D,B,A,A,A,A)
(C,D,D,C,D,D,C,D)	(A,B,C,D,D,C,C,C)	(C,D,C,A,B,B,B,B)
(D,D,C,D,D,C,D,C)		(B,A,B,C,D,D,D,D)
(D,C,D,D,C,D,C,D)		
(C,D,D,C,D,C,D,D)		

Table A-1: Strict Nash equilibria. The format is $(a_1, ..., a_8)$ where $a_i, i = 1, ..8$ is the action of player i.

B Appendix: Learning Rules: Algorithms

In this subsection we present the algorithms corresponding to each learning rule. In each round, subjects play a 4×4 game against their neighbors and the set of actions is $\{a, b, c, d\}$ for all players.

In reinforcement learning, subjects choose strategies that have performed well in the past with larger probabilities. Formally, at period t each subject i has a propensity to play each of her four actions. Let $q_i(z, t)$ represent subject i's propensity at time t of playing action z, for all t and $z \in \{a, b, c, d\}$. These propensities are updated by adding the payoff ϕ received in period t for playing action z to the previous propensity. Therefore, the updating rule is: $q_i(z, t+1) = q_i(z, t) + \phi$ if z was played in t and $q_i(z, t+1) = q_i(z, t)$ when i chose an action different from z in period t. Thus actions that achieved higher returns are reinforced and player i chooses action z at round t if

$$q_i(z, t) \in \max\{q_i(a, t), q_i(b, t), q_i(c, t), q_i(d, t)\} \tag{8}$$

The second class of learning model we consider is imitation learning model. Let N_i^R denote the set of Rth order neighbors of any subject i, with $R \in \{1, 2, ..., M\}$ and, cardinality n_i. In payoff based imitation order R, learners copy the most successful strategy within their Rth order neighbors. Let $\Delta_i^R(z, t)$ represent the average payoff of those players who played action z in round t within subject's i Rth order neighborhood. Player i, then, at time t chooses action z if

$$\Delta_i^R(z, t) \in \max\{\Delta_i^R(a, t), \Delta_i^R(b, t), \Delta_i^R(c, t), \Delta_i^R(d, t)\} \tag{9}$$

Under belief learning models subjects form beliefs on their opponents' strategies and choose an action that best responds to those beliefs. Let v_i be a vector whose elements, $v_i(z, t)$ represent the weight subject i gives to her opponents playing each pure strategy z in round t. Therefore player i believes her opponents in round t play action z with probability $p_i(z) = \frac{v_i(z,t)}{\sum_{s \in \{a,b,c,d\}} v_i(s,t)}$. Player i then chooses a pure strategy that is a best response to the probability distribution. A fictitious player consider the whole history of the game to compute her probability distribution. Let $Z_i(z, t)$ represent the set of player i's first order neighbors who played pure strategy z at round t with cardinality $n_i(z, t)$. At the first round no weight is put in any strategy, and hence, fictitious players choose randomly. For all subsequent periods a fictitious player updates her belief vector by $v_i(z,t) = v_i(z,t-1) + n_i(z,t)$. On the other hand, a myopic best responder only uses the most recent period to form her beliefs. Therefore, the updating rule for a myopic best responder is $v_i(z,t) = n_i(z,t)$.

Our last belief-based learning model is forward looking order 2 in which players assume their first order neighbors are myopic best responder and, consequently, choose a best response to their first order neighbors' myopic best response. Let $q(i,t)$ be a vector containing a number of elements equal to the number of player i's first order neighbors. Each element of $q(i,t)$ represents player i's first order neighbor's myopic best response at round t. Thus player i chooses a pure strategy that is a best response to $q(i,t)$.

For all learning rules, in case of tie, the player is assumed to choose randomly between the options that tie.

C Appendix: Additional Tables

Heterogeneous Network 2

	RL	PBI	MBR	FL
$N-2$ without costs	0.92	0.88	0.95	0.74
$N-2$ with costs	0.40	0.07	0.48	0.13

Table 14: Occurrence's discriminatory power. The percentage of rounds subjects consulted the minimal information required by each learning rule if information is costless (above), compared to $N-2$ (below).

C.1 Treatment N-1 Estimations with Information Search

	Treatment N-1 ($LL = -1857$)			
Parameters	RL	PBI	MBR	FL
p_k	0.21	0.62	0.17	0
θ_Z	0.03	0.99	0.05	-
θ_M	0	0	0	-
θ_H	0.97	0.01	0.95	-
ε_Z	1	1	1	-
ε_M	-	-	0.63	-
ε_H	0.56	0.11	0.31	-

Table 15: Information Search and Decisions *N-1,* all rules.

	Treatment N-1 ($LL = -3023$)			
Parameters	RL	PBI	MBR	FL
p_k	0.20		0.23	0.57
θ_Z	0.09		0.07	**0.99**
θ_M	0		0	0.01
θ_H	0.91		0.92	0
ε_Z	1		1	1
ε_M	-		-	0.99
ε_H	0.46		0.35	-

Table 16: Information Search and Decisions

C.2 Treatment N-2 Estimations with Information Search

	Treatment N-2 ($LL = -2762$)			
Parameters	RL	PBI	MBR	FL
p_k	0.29	0.36	0.23	0.12
θ_Z	0.04	**1**	0.03	0.98
θ_M	0	0	0	0.02
θ_H	0.96	0	0.97	0
ε_Z	1	1	1	1
ε_M	-	-	-	1
ε_H	0.45	-	0.55	-

Table 17: *Information Search and Decisions*

	Treatment N-2 ($LL = -4260$)			
Parameters	RL	PBI	MBR	FL
p_k	0.22		0.29	0.48
θ_Z	0.08		0.11	**0.98**
θ_M	0		0	0.02
θ_H	0.92		0.89	0
ε_Z	1		1	1
ε_M	-		-	0.77
ε_H	0.43		0.41	-

Table 18: *Information Search and Decisions*

C.3 Treatment N-3 Estimations with Information Search

	Treatment N-3 ($LL = -2010$)			
Parameters	RL	PBI	MBR	FL
p_k	0.06	0.62	0.30	0.03
θ_Z	0.15	**0.98**	0.03	0.49
θ_M	0.02	0.02	0	0.50
θ_H	0.83	0	0.97	0.01
ε_Z	0.41	1	0.92	0.51
ε_M	0.8	0.76	-	0.81
ε_H	0.38	-	0.61	0.76

Table 19: *Information Search and Decisions*

	Treatment N-3 ($LL = -2645$)			
Parameters	RL	PBI	MBR	FL
p_k	0.10		0.18	0.72
θ_Z	0.16		0.14	**0.97**
θ_M	0.02		0.01	0.03
θ_H	0.82		0.85	0
ε_Z	0.69		0.98	1
ε_M	0.62		0.39	0.59
ε_H	0.31		0.66	-

Table 20: *Information Search and Decisions*

C.4 Estimations with Variants of Fictitious Play

Treatment N-1 ($LL = 1410$)		
Parameters	RL	FP_3
p_k	0.58	0.42
θ_Z	0.56	0.10
θ_M	0.06	0
θ_H	0.38	0.90
ε_Z	1	1
ε_M	0.53	-
ε_H	0.51	0.47
Treatment N-2 ($LL = -2022$)		
Parameters	RL	FP_6
p_k	0.57	0.43
θ_Z	0.48	0.16
θ_M	0.09	0
θ_H	0.44	0.84
ε_Z	1	1
ε_M	0.47	-
ε_H	0.52	0.42
Treatment N-3 ($LL = -1479$)		
Parameters	RL	FP_6
p_k	0.67	0.33
θ_Z	0.86	0.66
θ_M	0.14	0.09
θ_H	0	0.25
ε_Z	1	1
ε_M	0.64	0.70
ε_H	-	0.70

Table 21: Variants of belief-based learning

C.5 Robustness Estimations

Table A-14

Treatment N-1 ($LL = -1081$)		
Parameters	RL	MBR
p_k	0.48	0.52
θ_Z	0.14	0.03
θ_M	0	0
θ_H	0.85	0.97
ε_Z	1	1
ε_M	-	-
ε_H	0.66	0.41
Treatment N-2 ($LL = -1408$)		
Parameters	RL	MBR
p_k	0.49	0.51
θ_Z	0.11	0.02
θ_M	0	0
θ_H	0.89	0.98
ε_Z	1	1
ε_M	-	-
ε_H	0.44	0.51
Treatment N-3 ($LL = -1179$)		
Parameters	RL	MBR
p_k	0.43	0.57
θ_Z	0.21	0.04
θ_M	0	0
θ_H	0.79	0.96
ε_Z	1	1
ε_M	-	-
ε_H	0.41	0.73

D Appendix: Sample Instructions (Treatments $N-1$, $N-2$ and $N-3$)

Welcome and thanks for participating at this experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules we are sorry to have to exclude you from the experiment. Please do also switch off your mobile phone at this moment.

For your participation you will receive 2 Euros. During the experiment you can earn more. How much depends on your behavior and the behavior of the other participants. During the experiment we will use ECU (Experimental Currency Units) and at the end we will pay you in Euros according to the exchange rate 1 Euro = 75 ECU. All your decisions will be treated confidentially.

THE EXPERIMENT

In the experiment you are linked up with some other participants in this room, which we will call your neighbors. You will play a game with your neighbors that we will describe below. Your neighbors in turn are of course linked up with you, but (possibly) also with other participants in the room. And their neighbors again are linked up with other participants and so on...

Note that your neighbors are not necessarily the participants who are located to your left and right in the physical layout of the computer laboratory.

During the experiment, you will be able to find out how many neighbors you have as well as their experimental identity, but not who they really are. This also means, of course, that your neighbors will not know your real identity.

The experiment lasts for 20 rounds. In each round you play a game with each of your neighbors. Your payoff in each round is the average payoffs obtained in all the games with your neighbors.

Each round consists of three stages, which we will describe in detail below. Here is a summary:

1. In the first stage you choose an action in the game. Note that you have to choose the same action for all your neighbors.

2. In the second stage you can request information about your neighbors, your neighbors' neighbors etc... the actions they chose in the past period and the payoff they obtained in the past period, as well as about your own payoff.

3. In the third stage, the information you requested is displayed on the computer screen.

We will now describe the different stages in more detail.

Stage 1 (Action Choice)

In the first stage you have to choose one action in the game, which is described by the following table, which will be shown to you every time you choose an action.

	A	B	C	D
A	20,20	40,70	10,60	20,30
B	70,40	10,10	30,30	10,30
C	60,10	30,30	10,10	30,40
D	30,20	30,10	40,30	20,20

In the table your actions and payoffs are given in dark grey and your neighbor's actions and payoffs in light grey. The table is read as follows (dark payoffs):

- If you choose A and your neighbor A, you receive 20
- If you choose A and your neighbor B, you receive 40
- If you choose A and your neighbour C, you receive 10
- If you choose A and your neighbor D, you receive 20
- If you choose B and your neighbor A, you receive 70
- If you choose B and your neighbor B, you receive 10
- If you choose B and your neighbor C, you receive 30
- If you choose B and your neighbor D, you receive 10
- If you choose C and your neighbor A, you receive 60
- If you choose C and your neighbor B, you receive 30
- If you choose C and your neighbor C, you receive 10
- If you choose C and your neighbor D, you receive 30
- If you choose D and your neighbor A, you receive 30
- If you choose D and your neighbor B, you receive 30
- If you choose D and your neighbor C, you receive 40
- If you choose D and your neighbor D, you receive 20

Note that your neighbor (light payoffs) is in the same situation as you are. This means that for your neighbor:

- If your neighbor chooses A and you A, your neighbor receives 20
- If your neighbor chooses A and you B, your neighbor receives 40
- If your neighbor chooses A and you C, your neighbor receives 10
- If your neighbor chooses A and you D, your neighbor receives 20
- If your neighbor chooses B and you A, your neighbor receives 70
- If your neighbor chooses B and you B, your neighbor receives 10
- If your neighbor chooses B and you C, your neighbor receives 30
- If your neighbor chooses B and you D, your neighbor receives 10
- If your neighbor chooses C and you A, your neighbor receives 60
- If your neighbor chooses C and you B, your neighbor receives 30
- If your neighbor chooses C and you C, your neighbor receives 10
- If your neighbor chooses C and you D, your neighbor receives 30
- If your neighbor chooses D and you A, your neighbor receives 30
- If your neighbor chooses D and you B, your neighbor receives 30
- If your neighbor chooses D and you C, your neighbor receives 40
- If your neighbor chooses D and you D, your neighbor receives 20

Remember that you have to choose the same action for all your neighbors. Your gross payoffs in each round are given by the sum of payoffs you have obtained in all games against your neighbors divided by the number of neighbors you have.

Stage 2 (Information Request)

In the second stage you can indicate which of the following pieces of information you would like to obtain

- the experimental identity of your neighbors

- the experimental identity of your neighbors' neighbors (2nd order neighbors)
- the experimental identity of your neighbors' neighbors' neighbors (3rd order)
- the experimental identity of your neighbors' neighbor's neighbors' neighbors (4th order neighbors)

Note that who is a neighbor of you does not change during the experiment. Hence once you have asked for this information in some round, it will be displayed in all future rounds. Note also that in order to receive information about your neighbors' neighbors' you first need to request information about your neighbors etc. The cost of requesting each of these pieces of information is 10. You only have to pay this cost once. In addition you can request information about the following items which (in principle) can change in every round.

- the actions chosen by your neighbors
- the actions chosen by your neighbors' neighbors
- the actions chosen by your neighbors' neighbors' neighbors
- the actions chosen by your neighbors' neighbor's neighbors' neighbors
- the payoffs obtained by your neighbors
- the payoffs obtained by your neighbors' neighbors
- the payoffs obtained by your neighbors' neighbors' neighbors
- the payoffs obtained by your neighbors' neighbor's neighbors' neighbors
- your own payoffs

Obviously, in order to receive information about your neighbors (or neighbors' neighbors') actions or payoffs you first need to request information about the experimental identity of your neighbors (neighbors' neighbors) etc. The cost of requesting each of these pieces of this information is 1 and you have to pay it each time you request this information anew. Your net payoffs in a round are your gross payoffs minus the cost of the information you requested.

Stage 3 (Information Display)

The information you have requested in Stage 2 is displayed on the screen for 40 seconds.

Control Questions

Before we start the experiment please answer the following control questions on your screen.

1. Assume you have only one neighbor. She chooses action B and you action D. Which gross payoff will you get in this round?
2. Assume you have three neighbors and they choose action A, B and A. You choose action D. Which gross payoff will you get in this round?
3. True or False: My neighbors change in every round of the game.
4. True or False: My neighbors face the same payoff table as I do.
5. True or False: My neighbors are those sitting in the cubicles to my left and right.