IKERLANAK

LEARNING IN NETWORK GAMES

by

Jaromír Kovárík, Friederike Mengel and José Gabriel Romero

2012

Working Paper Series: IL. 66/12

Departamento de Fundamentos del Análisis Económico I

Ekonomi Analisiaren Oinarriak I Saila

UPV EHU

University of the Basque Country

Learning in Network Games*

Jaromír Kovářík[†]
University of the Basque Country

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

José Gabriel Romero[¶]
University of Santiago de Chile

November 23, 2012

Abstract

We report the findings of an experiment designed to study how people learn and make decisions in network games. Network games offer new opportunities to identify learning rules, since on networks (compared to e.g. random matching) more rules differ in terms of their information requirements. 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' decisions are best characterized by reinforcement learning or (myopic) best-response learning. The distribution of learning types seems fairly stable across contexts. Neither network topology nor the position of a player in the network seem to substantially affect the estimated distribution of learning types.

JEL Classification: C72, C90, C91, D85.
Keywords: Experiments, Game Theory, Heterogeneity, Learning, Maximum Likelihood Method, Networks.

*We wish to thank Clayton Featherstone, Sanjeev Goyal, Nagore Iriberri, Matthew Jackson, Muriel Niederle, Aljaz Ule, Marco van der Leij and Nathaniel Wilcox as well as audiences in Cambridge, Chapman University, Navarra, Stanford, 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).

‡Department of Economics,University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom (fr.mengel@gmail.com).

§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

In many situations of economic interest people arrive at their decisions via a process of learning. For example, consider decisions such as how to conduct business negotiations, which projects to dedicate effort to, and in which assets to invest our money. Economists have developed a number of different models to describe how people learn in such situations (Fudenberg and Levine, 1998). These models, however, often lead to very different predictions. In a Cournot duopoly, for example, imitation learning can lead to the Walrasian outcome (Vega Redondo, 1997), while most belief learning models converge to the Cournot-Nash outcome. In a Prisoner's dilemma, some forms of aspiration based learning can lead to cooperation (Karandikar et al, 1998), while imitation and belief learning models will typically lead to defection. Hence to make predictions in these situations, it seems crucial to have some understanding about how people learn. In this paper we conduct an experiment designed to study how people learn in games, whereby we pay particular attention to (i) individual heterogeneity and (ii) to the question of whether learning rules people employ are stable across different contexts.

There is a considerable amount of existing research aimed at understanding how people learn in games. By far the most common method to study learning in experiments has been the representative-agent approach, where one 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) has shown that in the presence of heterogeneity estimating 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).Overall, this research has provided mixed evidence on which learning model best describes behaviour and models that have found support in some studies have been rejected in others.[1]

Other approaches include Camerer and Ho (2002), who assume that agents fall into two segments of subpopulations with different parameter values for each, or Camerer et al. (2002), who estimate a mixture of standard and sophisticated experience weighted attraction (EWA) learning in the population. Other studies have estimated 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]

Irrespective of whether a representative agent or an individual approach is used, many studies restrict the information feedback given to participants thereby ruling out some types of learning ex ante. If e.g. no information about payoffs of other participants is provided then e.g. payoff-based imitation learning can often be impossible.

In this paper we conduct an experiment where information about past play and payoffs of all participants as well as about network neighbours is available, but where we keep track of which information each participant requests between rounds. We combine this information with information about observed action choices to estimate a distribution of learning types using maximum likelihood methods. The advantage of observing both action choices and information requests is that even if different learning rules predict the same action choices, they can be distinguished as long as different

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

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

information is needed to reach that decision.[3]

We study strategic interaction in networks. Network games offer new opportunities to identify and separate different learning rules. The reason is that in networks (compared to random matching or fixed pairwise matching scenarios) it is more often possible to distinguish learning models via information requests. For example, consider myopic and forward-looking best reply learning. Under random matching an agent needs to know the distribution of play in the previous period irrespective of whether she is myopic or forward-looking. In a network, though, a myopic best responder needs to know *only* the past behavior of her first-order neighbours (who she interacts with), while a forward-looking learner may need to know the behavior of her second-order neighbours to be able to predict what her first-order neighbours will choose in the following period.[4] 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 predict different choices 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 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 study we assume that each player learns according to a certain learning rule, which is drawn from a common prior distribution, that we aim to estimate. In our baseline, we consider four prominent classes of learning models as possible descriptions of subjects' behavior. One of our learning types is reinforcement learning, another is imitation learning, and two models are belief-based (myopic best response and forward-looking learning).

In total our experiment consists of six treatments. Three treatments with endogenous information request (for three different network topologies) and three control treatments with the same networks but *without* endogenous information request. 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 existence of information requests *per se* affects action choices and whether participants in the endogenous information treatments request all the information they would naturally use in making their decisions. We find no significant differences between action choices in the control treatments and the treatments with endogenous information request.

We now briefly summarize our main results. There is substantial heterogeneity in the way people learn in our data. However, most of our participants' decisions are best described by either reinforcement learning or myopic best response learning. There is little evidence of forward looking behaviour. Even though we observe significant effects of the network topology on participants' action choices,

[3]Indeed, Salmon (2001) has shown that estimations of learning models based only on choice behaviour can easily fail to detect the underlying data generating process.

[4]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.

our results suggest that network topology has limited influence on how people learn. The estimated distribution of learning types is virtually identical in two of our more homogeneous networks. In our (star-like) network with the highest variance in the degree distribution the estimated distribution of learning types gives more weight to belief learning rules at the expense of reinforcement learning.

We also ask whether there are differences in learning across participants with different network positions. In particular we ask whether having fewer or more network neighbours and whether being in a more central or peripheral position in the network affect learning. Again, we find no substantial differences. The estimated distribution of learning types is remarkably stable across players with many and few neighbours and across central and peripheral players.

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 requests. 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 choice 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 our experimental treatments participants repeatedly played the symmetric two player game depicted in Table 1 with their (first-order) neighbours in the network. Within each session the networks were fixed, which means that each participant played with the same neighbours in all of 20 periods. Each player had to choose the *same* action against all her neighbours. If participants were allowed to choose different actions for their different neighbours, 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 (bilateral) games against the neighbours. We chose to pay the average rather than the sum of 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. We chose a 4×4 rather than a 2×2 game, because (i) we hoped

	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	$30, 20$	$30, 10$	$40, 30$	$20, 20$

Table 1: The One-Shot Game.

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 easier to identify a larger number of different learning rules from observing agents' choices only. Hence, by choosing this game we hoped to give good chances to estimations based on choice behaviour alone and as a consequence have a stronger case for estimations including information requests if they prove to give superior estimates.

The treatments differed along two dimensions: network architecture and information accessibility. Throughout the paper we denote network architectures by numbers 1, 2 and 3 (see Figures 1-3) and information levels by 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

In our experiment, we used three different networks (one very homogenous, one very heterogeneous and one intermediate) to see whether people might resort to different learning rules in different environments. Figures 1-3 show the three network architectures used in the experiment and Table 2 summarizes the most standard network characteristics of these networks. 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 denotes the degree of agent i (i.e. the number of i's first-order neighbours) 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 other network characteristics are kept approximately constant. We selected networks with different variances $\sigma^2(\kappa)$ to see whether and how choice behaviour or learning are affected if there are strong asymmetries in the environment.

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 neighbours choose. For example in Network 1 (see Figure 1) one network equilibrium is that players 1,..,8 choose actions $(a_1, .., a_8) = (A, B, A, B, A, B, A, B)$. All players choosing A in this equilibrium get an average payoff of 40 (because both their neighbours choose B) and all players choosing B get a payoff of 70 (because both their neighbours choose A).

In the following, whenever we refer to equilibria we will refer to such network equilibria. All our networks are designed such that many strict pure strategy equilibria exist in the one-shot network game (between 9-12 depending on the network). 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.

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) action choices corresponded even approximately to a Nash equilibrium of the repeated game. Choices did converge, though, to a one-shot network equilibrium in several networks (see below).

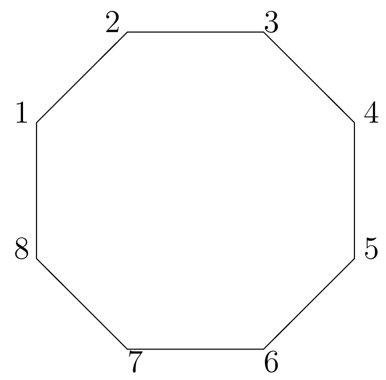


Figure 1: Treatments N-1 and F-1

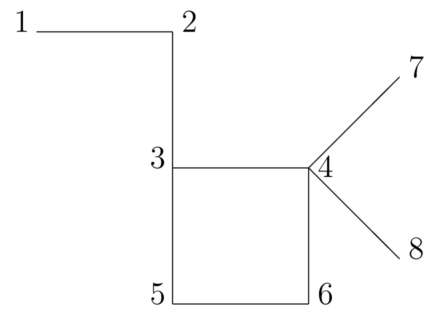


Figure 2: Treatments N-2 and F-2

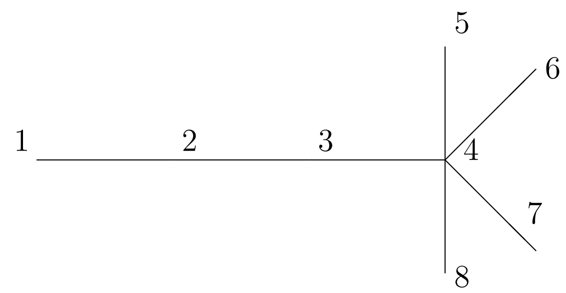


Figure 3: Treatments N-3 and F-3

2.3 Information

Our second treatment dimension varies 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. Hence participants did not know what the network looks like or what actions other players chose etc. Instead, at the beginning

	$N-1$	$N-2$	$N-3$
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.

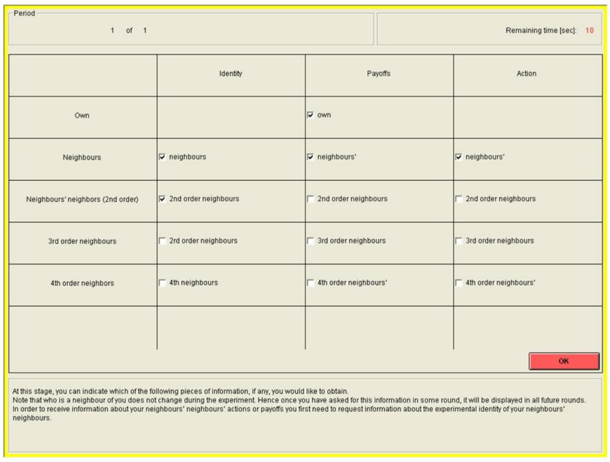
of each round they were asked which information they would like to request. They could request three types of information: (i) the network structure, (ii) past action choices and (iii) past payoffs of their network neighbours. More precisely, if a participant requests the network position of her first order neighbours she is shown how many neighbours she has and their experimental label (which is a number between 1 and 8; see Figures 1-3). With second- order neighbours, she is shown their experimental label as well as the links between the first and second-order neighbours. For third and fourth-order neighbours she is shown the analogous information. Regarding actions and payoffs, subjects are shown the actions and/or payoffs of their first-, second-, third- and/or fourth-order neighbours if they requested 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 requests. First, subjects were only allowed to ask for the actions and/or payoffs of subjects whose experimental label they had previously requested. Second, they were not allowed to request the experimental label of higher order neighbours without knowing the label of lower order neighbours. 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.

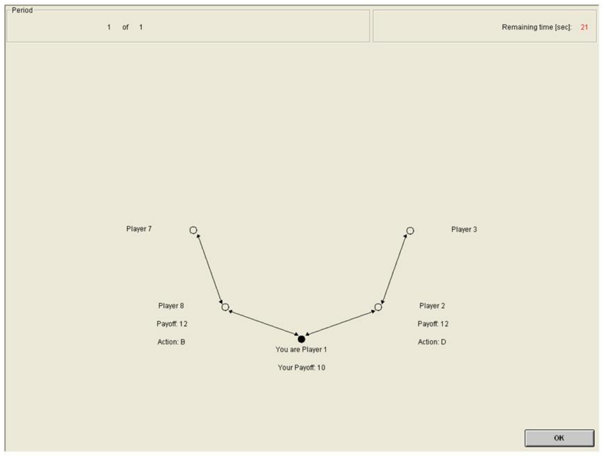
Imposing a (small) cost on information requests is a crucial element of our design. Of course, even though costs are "small" this does affect incentives. We imposed costs to avoid that participants request information they are not using to make their choices. We also conducted one treatment that coincided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

To see whether information requests per se affect action choices (e.g. because participants might not look up "enough "information due to the costs) we conducted three control treatments with full information. Those treatments $F-1$, $F-2$ and $F-3$ coincided with N-1, N-2 and N-3, respectively, but there was no information request stage. Instead, all the information was displayed at the end

[5]An alternative approach was taken by CCB. They use the computer interface MouseLab to monitor mouse movements. 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.



(a) Screen: Information Requests



(b) Screen: Information Display

Figure 4: Screenshots of the screen for information requests and information display.

of each period to all participants. We call these the *full information* treatments. We can use those treatments to study whether there are any differences in choice behavior induced by the existence of costly information requests. 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 the BEE-Lab 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 descriptive statistics regarding both action choices and information requests.

3.1 Choice Behaviour

Network topology does impact action choices. First, in terms of convergence to a Nash equilibrium of the network game, we observe relatively high convergence in Network 2, intermediate levels in

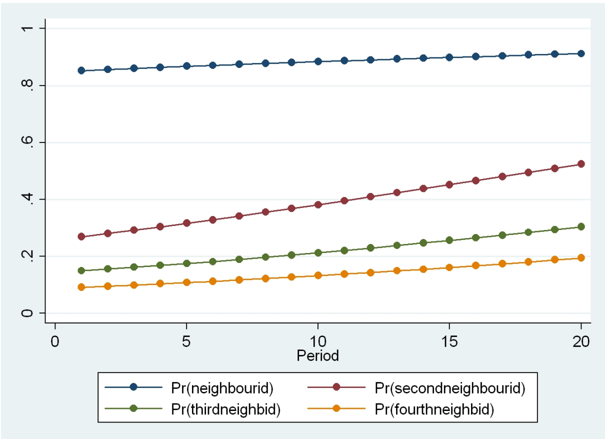
[6]The table does not contain the treatment $N-2$ without costs mentioned above. We will not discuss this treatment any further, but results are of course available upon request. Other than the treatments reported we didn't conduct any other treatments or sessions and we did not run any "pilot studies".

Network 1, followed by Network 3. The entire network converges to an equilibrium 12% (28% and 0%) of the time in the last 5 periods of play in $N-1$ ($N-2$ and $N-3$, respectively). All the differences are statistically significant if we take each network as an independent unit of observation (Mann-Whitney, $p < 0.01$). Convergence rates are not statistically different, however, along the information dimension (comparing F-1 with N-1, F-2 with N-2 and F-3 with N-3). In all treatments participants eventually coordinated on a network equilibrium where all players choose either C or D (if they coordinate at all). Full and Endogeneous Information treatments are statistically no different in terms of which equilibrium play converges to nor in terms of the overall distribution of action choices in the 20 periods. (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) did *not* seem to distort the choice behaviour of our experimental subjects.

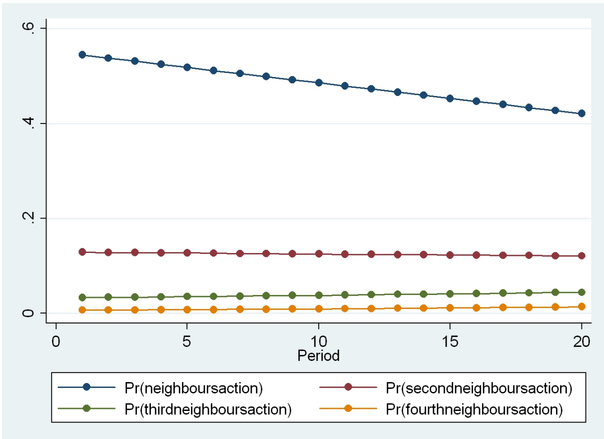
3.2 Information Requests

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

Network Structure. Figure 5 (left panel) illustrates information requests concerning the network structure, aggregated over treatments. In the first round 77.5%, 76.8% and 72.5% of participants in $N-1$, $N-2$ and $N-3$, respectively, requested the label of their direct neighbours. Roughly 90% of individuals end up requesting 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 neighbours (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.



(a) Information Requests Labels



(b) Information Requests Action Choices

Figure 5: Information Requests. Note the different scale of the y-axis.

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

Actions. Around 50% of participants request information about past actions of their opponents (i.e. their direct neighbours 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 neighbours' 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 request information about choices of third- or fourth-order neighbours. Figure 5 (right panel) shows the evolution of the action-related information requests over time, aggregated over all the treatments.

4 Framework

This section discusses our selection of learning models and sets out basic issues in identifying learning rules from our data (i.e. action choices and information requests). In our baseline specification, 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 can be found in the Appendix:

1. Under *Reinforcement Learning* (*RL*) participants randomize in each period between actions with probabilities that are proportional to past payoffs obtained with these actions (Roth and Erev, 1995; Erev and Roth, 1998; Sutton and Barto, 1998; Skyrms and Pemantle, 2000; Hopkins, 2002).[7]

2. Under *Payoff-based Imitation* (*PBI*) participants choose the most successful action from the previous period (i.e. the action with the highest average payoff) in an agent's first-order neighbourhood including the agent herself (Eshel, Samuelson and Shaked, 1998; Bjoernerstedt and Weibull, 1995; Vega Redondo, 1996, 1998; Schlag, 1998; Skyrms and Pemantle, 2000; Alos-Ferrer and Weidenholzer, 2008; Fosco and Mengel, 2011).[8]

3. Under *Myopic Best Responses* (*MBR*) players choose a myopic best-response to the choices of their first-order neighbours in the previous period (Ellison, 1993; Jackson and Watts, 2002; Goyal and Vega Redondo, 2005; Hojman and Szeidl, 2006; Blume, 1993).

4. *Forward-Looking (FL)* subjects in our specification assume that their first-order neighbours are myopic best responders and best-respond to what they anticipate their first-order neighbours to play in the following period (Blume, 2004; Fujiwara-Greve, 1999; Mengel, 2012).

In Section 6 we also include some variants of these rules, such as fictitious play learning (with different memory lengths). In several robustness checks we also included less well-known rules such as conformist imitation rules, aspiration based reinforcement learning rules and variants of the payoff-based imitation rule. These rules only differ from the above rules in very few instances of predicted action choices, but not in terms of information requests (results are available upon request). The four rules singled out above are each representative of a larger class of learning models. Including all possible variants would mean to over-specify the model considerably.

[7] In fact in our estimations we will assume that a participant perfectly consistent with RL chooses the most preferred action with probability one (see below). This approximates some exponential choice rules used in the literature, but is not the case with e.g. the linearly proportional rule.

[8] Some of these authors study, in fact, imitation of the action with the maximal payoff obtained by any single agent instead of the highest average payoff. Using this variation does not fundamentally alter any of our results.

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. The main reason is that level-k learning - despite its name - is a model of initial responses and 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.

4.1 Identifying Learning Rules from Observed Action Choices

Obviously, it is only possible to identify learning rules if different rules predict different choices and/or information requests in the experiment. Table 4 presents the average number of rounds in which two different learning types predict different choices for a participant in our experiment.

In all treatments, the number of rounds RL prescribes different action choices 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 rounds). The number of rounds in which choices implied by PBI are different from those of belief learning ranges from 11 to 15 rounds. Finally, the number of rounds in which choices implied by MBR and FL 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 action choices 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

Table 4: Separation between learning types on basis of action choices. Each cell contains the average number of rounds in which the two corresponding types predict different choices.

4.2 Identifying Learning Rules from Information Requests

Apart from choices we also observe participants' information requests. As mentioned above, our different learning rules imply different needs in terms of information. A reinforcement learner only needs to consult her own past payoffs. Payoff-based imitation requires participants to consult their first-order neighbours' identities, actions, and payoffs, while a myopic best responder has to consult her first-order neighbours identities and action choices. Forward-looking learners should request their first- and second-order neighbours' identities and action choices. Table 5 summarizes this information. If participants always requested exactly the minimal information needed for a learning type, then all rules could be identified in all of the 20 periods.

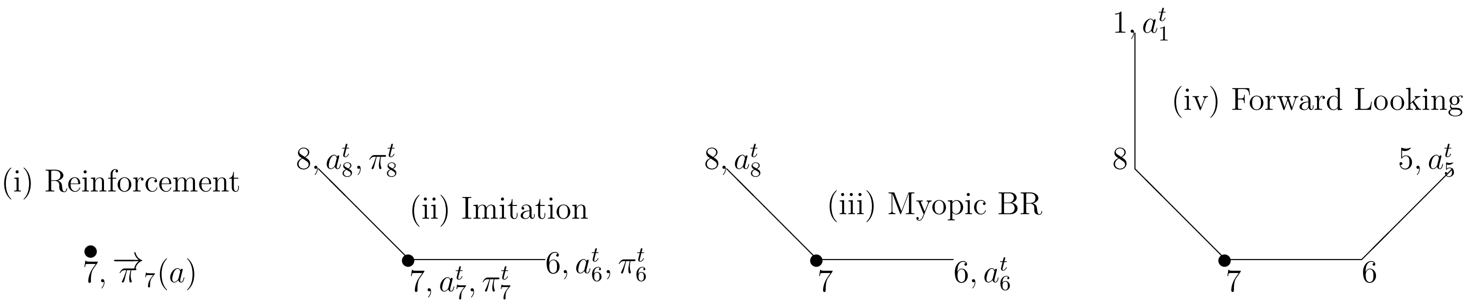


Figure 6: Player 7 in Network 1. Information required for Rules (i) RL, (ii) PBI, (iii) MBR and (iv) FL. a_i^t denotes the action taken by player i at time t. π_i^t the payoffs obtained by player i at time t and $\pi_i(a)$ the vector of average payoffs obtained by player i with each of the four actions.

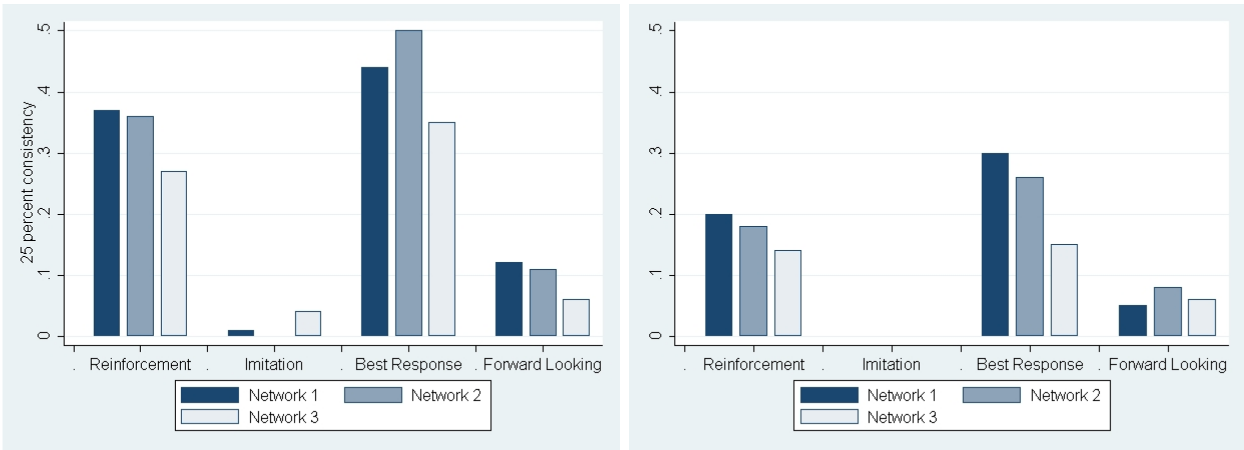
		Learning Type			
Info	Neighbour	*RL*	*PBI*	*MBR*	*FL*
Id	1		x	x	x
	2				x
Action	1		x	x	
	2				x
Payoff	Own	x	x		
	1		x		

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

Figure 6 provides an example of how different learning rules imply different information requests. The example focuses on player 7 in Network 1. As a reinforcement learner she doesn't need any knowledge of the network or the choices of others. In fact she wouldn't even need to know the payoff matrix. The only knowledge she needs is her average payoffs in the past with each of the actions. Using payoff based imitation she would need information about the choices of her first-order neighbours and the payoffs they obtain. Under myopic best responses she would need to know choices of her first order neighbours and as a forward looking agent she would also need to know her second order neighbours and choices. Note that under common knowledge of rationality and correct beliefs (the assumptions underlying Nash equilibrium), strategic decision makers would not need to request any information about past action choices nor about past payoffs of themselves or their neighbours.

An important 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 neighbours and then uses this information together with the payoff matrix to compute the payoffs of her first-order neighbours. Clearly, we cannot avoid this. Our design is such, however, that it is always *more* costly (in terms of the costs we impose on information requests) to make indirect inference about desired information rather than consulting it directly. This is hence an additional advantage of having small costs for information requests (in addition to those mentioned in Section 2).

4.3 Information Requests and Action Choices in the Data



(a) 25 percent consistency (b) 50 percent consistency

Figure 7: Fraction of participants who request the minimal information set *and* play the action prescribed by the corresponding learning type more than 25% (left) or 50% (right) of rounds.

In this subsection, we provide a descriptive overview of the performance of the different learning rules in the data. Figure 7 shows the fractions of individuals who *both* 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 are consistent with forward-looking learning, while most individuals' information requests are consistent with reinforcement learning or myopic best responses.

5 Maximum-Likelihood Estimations

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

5.1 Econometric Framework

We start with the following assumption, which links information requests 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.

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 Sections 2 and 4.2).

For each subject i and learning type $k \in \{1, 2, ..., K\}$, we then 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 the advantage that 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.

Let θ_{kj} denote the probability that a participant has compliance j with rule k in the experiment, where $j \in \{Z, M, H\}$ and $\sum_j \theta_{kj} = 1$. 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 $c \in \{1, 2, 3, 4\}$ denote the number of possible action choices predicted by a learning rule in a given round. 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 error rates are i.i.d 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 for a decision maker of type k to choose an action consistent with rule k (either by mistake or as a result of employing rule k) is

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

The probability to choose a given action that is inconsistent with rule k is $\frac{\epsilon_{kj}}{4}$. For each learning rule k in each period we observe which action a player chooses and whether or not it is consistent with learning rule 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 when participant i is of type 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). 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, $2K$ information request probabilities θ_{kj}, and $3K$ error rates ε_{kj}. It is well known that it is easy to overparameterize this sort of finite mixture models and standard information criteria for model selection, such as the Aikaike or Bayesian Information Criteria, might not perform satisfactorily (Prasad et al., 2007, Cameron and Trivedi, 2010). Below we describe how we proceed with model selection (i.e. selection of components) in our case.

First note that 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 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. A high concentration at zero compliance, for example, can lead to a probability θ_{kZ} very close to 1, and to a high estimated frequency p_k. However, a high 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. Finally, note that - given the shape of (3) - it is not feasible to estimate standard errors.

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 use the following three-step procedure.

Step (a) Estimate the model using all learning rules that have not been eliminated yet. (Initially this will be all four rules).

Step (b) If there is a learning type l with an estimated θ_{lZ} larger than the elimination threshold $\overline{\theta_Z}$, go to step (c). If every learning type has an estimated θ_{lZ} smaller than $\overline{\theta_Z}$ (i.e. if for every learning type the minimal information set was requested *at least* with probability $1-\overline{\theta_Z}$) and if estimated error rates increase as compliance decreases, stop the estimation process.

Step (c) If type l is the only type with $\theta_{lZ} > \overline{\theta_Z}$, eliminate type l from the set of rules considered. If there are multiple types with $\theta_{lZ} > \overline{\theta_Z}$, eliminate the type with the highest value of θ_{lZ}. Then go back to step (a).

The elimination threshold $\overline{\theta_Z}$ can in principle be set to any level depending on when one starts to believe that a rule fails to explain behaviour. Below we will consider values of $\overline{\theta_Z} \in [0.56, 0.97]$. The rules we eliminate usually have an estimated $\theta_{lZ} > 0.97$. In fact, any number between 0.56 and 0.89 as a threshold yields the same results in all networks. For a threshold between 0.90 and 0.97 we still get the same results in Networks 1 and 2 and in Network 3 reinforcement learning survives, while it is eliminated for $\overline{\theta_Z} < 0.9$.

5.2 Estimation Results with Information Requests and Choices

We start by illustrating how our algorithm selects learning rules. Table 6 shows the results for treatment $N-1$. The tables corresponding to treatments $N-2$ and $N-3$ can be found in the Appendix.

Table 6 shows the estimated type frequencies p_k and parameters θ_{kZ}. After initial 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' choice behavior was induced by the PBI learning rule and we remove PBI from the estimation. In the second iteration of the algorithm we eliminate the forward looking rule also with $\theta_{FWL,Z} = 0.99$. The algorithm stops with only two rules, RL and MBR, remaining. Our selection algorithm selects the same learning rules in $N-2$ for all $\overline{\theta_Z} \in [0.56, 0.97]$ and in treatment $N-3$ for all $\overline{\theta_Z} \in [0.56, 0.89]$ (see Tables 17-19 in the Appendix). We describe the result in more detail below.

How can it be that at the first step of estimations a rule that clearly doesn't describe behaviour well (like the rule PBI above) obtains an estimated value of p_{PBI} of 0.62? Remember that the estimation procedure identifies correlations between information requests and consistent choices. Hence if participants' choices do very poorly at explaining the variation in information requests, this will lead to a high concentration on zero compliance (Z) and will favor the estimated value of p. For this reason any estimated value of p_k can only be interpreted jointly with the vector θ_k (see Section 5.1).

There is additional information that can be gained by studying Tables 6 and 17-19. In $N-3$, for example, our population is overall best described by MBR and RL. 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 (second iteration, 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 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.

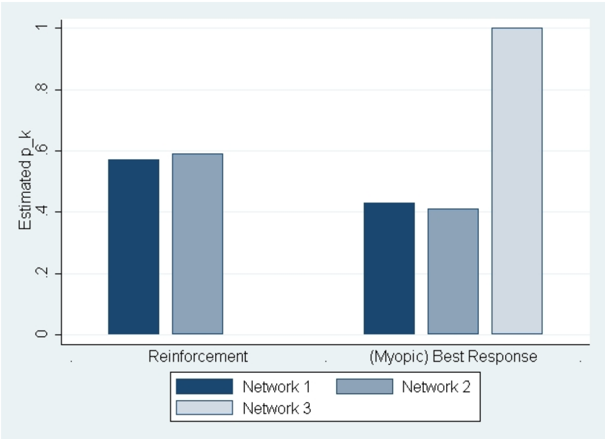
Table 7 reports the maximum likelihood estimates of learning type probabilities, p_k, unconditional compliance probabilities, θ_{kj}, and compliance conditional error rates, ε_{kj}, in the selected models. Figure 8 illustrates their estimated frequencies.

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.

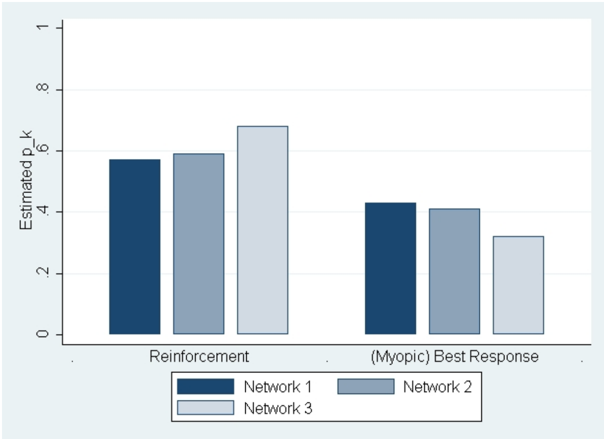
In $N-2$, 59% and 41% of subjects are best described by RL and MBR, respectively. The estimated θ's and ε's are also well behaved. Note that the estimates are remarkably similar in $N-1$ and $N-2$. In both networks, a combination of reinforcement learners and myopic best responders best describes the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

	Learning types			
Parameters	*RL*	*PBI*	*MBR*	*FL*
first iteration				
p_k	0.21	0.62	0.17	0
$\theta_{k,Z}$	0.03	**0.99**	0.05	-
second iteration				
p_k	0.20		0.23	0.57
$\theta_{k,Z}$	0.09		0.07	**0.99**
final estimation				
p_k	0.57		0.43	
$\theta_{k,Z}$	0.55		0.10	

Table 6: Selection Algorithm. Treatment N-1



(a) Thresholds between 0.56 and 0.89



(b) Thresholds between 0.90 and 0.97

Figure 8: Estimation of p_k using information requests and action choices.

Finally, for thresholds below 90% only MBR survives in $N-3$. In the final estimation 68% of subjects are classified as reinforcement learners, but they request information consistent with reinforcement learning (i.e. their own payoff) only with probability 0.1.

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 of subjects best described as reinforcement learners and the remaining participants as belief-based learners. MBR seems to play a more important role in $N-3$.

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 ever 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

	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 request and Decisions

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 requests and "correct" choices given the learning models consistent with the information request. As a result, if a decision-maker always looks up some information for other reasons (unrelated to the way she learns and plays), 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 neighbours in an Anti-Coordination game. To address this issue, we verify whether imitation of second-order neighbours (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 neighbourhood.

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 request. Allowing for unlikely or unpopular rules also has the advantage that we can check whether our estimation procedure correctly identifies them as such. 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 this issue for future research.

5.3 Estimation Results without Information Requests

How important is the role of monitoring information requests in our estimations? To address this question, in this section we estimate learning types solely on basis of participants' observed choices (disregarding their information requests). The objective is to show that if we only use information about subjects' action choices the estimates are less accurate, despite 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 any 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, meaning that learning type k's decisions are 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 standard errors.

Table 8 reports estimated frequencies and error rates for each treatment. In all cases we have evidence in favor of all 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 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 roughly $15-30\%$ of participants are mis-classified if we only consider action choices and ignore information requests.

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. One would hence expect these biases to be much more severe for smaller games or pure Coordination games, where identification based on choices alone is more difficult.

How do we know that the model with information requests gives "better" and not just "different" estimates than the model without information request? Obviously estimations that take into account information requests 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 request and "correct" choices given the learning models consistent with the information requests. Hence if a decision-maker always requests 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 requests. The only case in which the process with information request 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

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.

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 requests even more strongly.

6 Further Results

In this section, we further address the question of stability across contexts and we also do some robustness estimations. 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 learn is affected by their position in the network. One might conjecture, for example, that players with more complex decision problems, such as e.g. those with many network neighbours might resort to simpler decision rules. 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 different categories according to two types of characteristics.

First, we aggregate data from the heterogeneous networks 2 and 3 and categorize people into two groups according to whether they have one neighbour or more than one neighbour. Group 1 (with one network neighbour) contains players 1,7,8 in N-2 and 1,5,6,7,8 in N-3 and group 2 (multiple network neighbours) contains players 2,3,4,5,6 in N-2 and 2,3,4 in N-3. Omitting N-1 gives us more balanced groups (8 player positions per group), which enables a better comparison between the two sets of estimations. Table 9 shows the results of the final estimation (after applying the elimination algorithm twice and eliminating Imitation and forward looking learning). The full sequence of estimations can be found in the Appendix in Tables 20-21. Those results show very few differences between the two sets of estimations. In the final estimation reinforcement learning gets somewhat more than 50 percent of the weight and belief learning somewhat less than 50 percent. Note, though, that θ_Z is relatively high in these regressions. Table 21 in the Appendix shows that in both cases about 17-18 percent of decisions (info requests and action choices) are very accurately described by reinforcement learning and 26 percent by myopic best response learning.

	One Network Neighbour LL=-1695.09				Multiple Network Neighbours LL=-1780.44			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.57	-	0.43	-	0.53	-	0.47	-
θ_Z	0.85	-	0.75	-	0.88	-	0.50	-
θ_M	0.15	-	0.13	-	0.12	-	0.08	-
θ_H	0	-	0.12	-	0	-	0.43	-
ε_Z	1	-	1	-	1	-	1	-
ε_M	0.64	-	0.81	-	0.42	-	0.53	-
ε_H	-	-	0.59	-	0	-	0.27	

Table 9: Participants with one network neighbour (1,7,8 in N-2 and 1,5,6,7,8 in N-3) vs participants with multiple network neighbours (2,3,4,5,6 in N-2 and 2,3,4 in N-3).

As a second exercise, we split participants within each network N-2 and N-3 according to whether they have a central or peripheral position. In N-2, players 1,2,7 and 8 are classified as "peripheral" and players 3,4,5 and 6 as "central", while in N-3 we classify players 1,5,6,7 and 8 as "peripheral" and players 2,3 and 4 as "central". Note that the correlation between "peripheral" and having only one neighbour is (almost) perfect. Hence, if we aggregated again the data from N-2 and N-3 we would essentially be running the same estimations again. As a consequence, we conducted these estimations separately for each network. Note, that this implies that we will have somewhat small samples in each group, which should be kept in mind when interpreting these results. Again we find very few differences between peripheral and central positions. In N-2, reinforcement learning very accurately describes around 18 percent of decisions by peripheral players and myopic best response learning about 30 percent (see Table 23). The more noisy behavior is attributed mostly, but not uniquely, to reinforcement learning, as a result of which θ_Z is quite high for both rules in the final estimation (Table 24). For central players reinforcement accurately describes around 24 percent of decisions and MBR around 31 percent (Table 23). The more noisy behavior here, is mostly attributed to (myopic) best response learning and θ_Z remains lower in the final estimation (Table 24). The picture looks very similar for peripheral players in N-3, while it is for the three central players that myopic best response seems a better description of behaviour. The tables with the full sequence of estimations for all these cases can be found in the Appendix (Tables 22-27). Overall, we conclude from this subsection that the estimated distribution of learning types is remarkably stable across different types of player positions.

6.2 Fictitious Play with Limited Recall

In this section we estimate model (3) with different variations of belief learning. In particular, we assume that subjects form beliefs based on a fixed number of past periods. Myopic best responders are at one end of this classification basing their decisions on the last round only. We consider six alternative specifications, where players form beliefs based on choices of their opponents in the last three, six, nine, twelve, fifteen and twenty past periods to construct their beliefs. Note that the last variation corresponds to standard fictitious-play learning in our context. Denote by FP_s the variation, under which subjects form beliefs based on the last s periods. Hence, under this terminology, the myopic best-response rule is denoted FP_1, and fictitious play corresponds 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 28 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 virtually 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 model containing MBR. In all cases FP_{20} is among the last in the ranking.

These results show that including fictitious play (or variants of it) instead of myopic best response learning does not significantly alter any of the results. They also show that belief based models focused on few past periods tend to explain data better than those based on many periods.

6.3 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 model assuming that the data-generating process (information requests and choices) 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 a learning rule has an estimated μ_k higher than a threshold $\overline{\mu}$ we remove it from the set of rules considered. Any threshold $\overline{\mu} \in [0, 1]$ yields the same results. 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 request 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_{RL} and λ_{RL} 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 our selection criterion all subjects are again classified as MBR learners as under (3).

Hence, this alternative 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.

6.4 Recovering the data generating process from simulated data

In this section we put our model through another test. We simulate data from a distribution of learning types and then see whether our estimations are able to recover the original data generating process. More precisely, 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. We conducted those simulations only for treatment N-1.

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 that those types are randomly distributed on the network. (This assumption is less problematic in treatment N-1, where all players have the same network position. In other networks the learning types can be endogenous to the network position as we have seen above.) We consider three different parameter constellations (summarized in Table 11) as follows:

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 request with high probability and make mistakes 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.

Table 12 reports the results of our estimations of the data generated in this manner. In all three cases our selection algorithm correctly identifies the learning rules present in the population. The shares of PBI and FL are virtually zero in all cases. Moreover, we find only small biases in

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

	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)	0.070	0.008	0.075	0.006
Bias	0.020	0.002	-0.030	0.001
HC				
$\widehat{p}_k$	0.571	0.000	0.428	0.000
(se)	0.044	0.004	0.043	0.000
Bias	-0.003	0.004	0.003	0.000
LC				
$\widehat{p}_k$	0.443	0	0.556	0
(se)	0.030	0	0.030	0
Bias	-0.132	0	0.132	0

Table 12: Monte Carlo Simulations.

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 in recovering the population composition in all cases.

As subjects become less precise in their information requests 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. Hence, overall these results make us confident that our estimation procedure works well.

6.5 Relaxing the Assumption of Occurrence

As a final robustness check we relax our assumption of occurrence. In particular, we will be less restrictive on how often participants should look up the minimal information required by each rule. This is important here, since we have observed that subject consulted the information less frequently after convergence to an equilibrium (Section 3).

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 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: Estimation Results under relaxed assumptions on Occurence.

7 Concluding Remarks

We estimated learning types of participants in network games via maximum likelihood methods. Our estimates are based on both knowledge about what participants choose and which information they request in order to make their choice. 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 learning rules people resort to are remarkably stable across contexts. 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 S. Weidenholzer (2008), Contagion and Efficiency, Journal of Economic Theory 143, 251-274.

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

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

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

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

[6] 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.

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

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

[9] 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.

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

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

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

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

[14] 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.

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

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

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

[18] Fosco, C. and F. Mengel (2011), Cooperation through Imitation and Exclusion in Networks, Journal of Economic Dynamics and Control 35, 641-658.

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

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

[21] 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.

[22] Galeotti, A., Goyal, S., Jackson, M., Vega-Redondo, F., and Yariv, L. (2010), Network games, Review of Economic Studies, 77(1), 218-244.

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

[24] 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.

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

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

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

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

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

[30] 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.

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

[32] Jackson, M. O. (2010), Social and Economic Networks, Princeton University Press.

[33] Jackson, M.O. (2011), An Overview of Social Networks and Economic Applications, in: Handbook of Social Economics, Benhabib, J. Bisin, A., and Jackson, M. eds.

[34] 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.

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

[36] Kirchkamp, O. and Nagel, R. (2007), "Naive learning and cooperation in network experiments", Games and Economic Behavior 58, 269-292.

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

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

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

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

[41] 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.

[42] 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.

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

[44] 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.

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

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

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

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

[49] 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 neighbours 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 neighbours 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 neighbours. 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 neighbourhood. 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 neighbours 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 learning model is forward looking learning in which players assume their first order neighbours are myopic best responder and, consequently, choose a best response to their first order neighbours' 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 neighbours. Each element of $q(i,t)$ represents player i's first order neighbour'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

C.1 Treatment N-1 Estimations with Information Requests

	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 14: 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 15: Information Search and Decisions

C.2 Treatment N-2 Estimations with Information Requests

	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 16: *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 17: *Information Search and Decisions*

C.3 Treatment N-3 Estimations with Information Requests

	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 18: *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 19: *Information Search and Decisions*

C.4 Estimations by Player Position

	One Network Neighbour LL=-2073.63				Multiple Network Neighbours LL=-2543.26			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.21	0.77	0.03	0	0.14	0.29	0.19	0.38
θ_Z	0.11	**0.98**	0.91	0.25	0.05	0.99	0.03	**1**
θ_M	0	0.02	0.08	0.27	0	0.01	0	0
θ_H	0.89	0	0.01	0.48	0.95	0	0.97	0
ε_Z	1	1	0.45	0.82	1	1	1	1
ε_M	0.09	1	0.53	0.41	0.52	1	0.90	1
ε_H	0.53	0.52	0.89	0.28	0.33	0.16	0.35	0.39

Table 20: First estimation. Participants with one network neighbour (1,7,8 in N-2 and 1,5,6,7,8 in N-3) vs participants with multiple network neighbours (2,3,4,5,6 in N-2 and 2,3,4 in N-3).

	One Network Neighbour LL=-3770.57				Multiple Network Neighbours LL=-4914.28			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.18	-	0.26	0.56	0.17	0.57	0.26	-
θ_Z	0.10	-	0.12	**0.97**	0.05	**0.98**	0.09	-
θ_M	0	-	0	0.02	0	0.02	0	-
θ_H	0.90	-	0.88	0	0.95	0	0.91	-
ε_Z	1	-	1	1	1	1	1	-
ε_M	0.90	-	0.60	0.73	0.31	0.79	0.11	-
ε_H	0.51	-	0.63	0	0.40	0.58	0.31	-

Table 21: Second estimation. Participants with one network neighbour (1,7,8 in N-2 and 1,5,6,7,8 in N-3) vs participants with multiple network neighbours (2,3,4,5,6 in N-2 and 2,3,4 in N-3).

.

	Periphery LL=-1306.79				Centre LL=-1430.35			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.11	0.61	0.28	0	0.28	0.24	0.20	0.28
θ_Z	0.04	0.97	0.08	0.36	0.05	0.99	0.02	0.99
θ_M	0	0.03	0	0.37	0	0.01	0	0
θ_H	0.96	0	0.92	0.28	0.95	0	0.98	0
ε_Z	1	1	1	1	1	1	1	1
ε_M	-	0.92	-	0.23	-	1	-	1
ε_H	0.44	0.50	0.42	0.01	0.36	-	0.45	0.42

Table 22: Network 2. Peripheral vs Central Players. First Estimation.

	Periphery LL=-2256.20				Centre LL=-2538.39			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.18	-	0.30	0.52	0.24	0.45	0.31	-
θ_Z	0.04	-	0.10	0.98	0.05	0.97	0.09	-
θ_M	0	-	0	0.02	0	0.03	0	-
θ_H	0.96	-	0.90	0	0.95	0	0.91	-
ε_Z	1	-	1	1	1	1	1	-
ε_M	-	-	-	0.71	-	0.60	0.58	-
ε_H	0.52	-	0.48	0.75	0.45	0.93	-	-

Table 23: Network 2. Peripheral vs Central Players. Second Estimation.

	Periphery LL=-987.91				Centre LL=-958.18			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.61	-	0.39	-	0.36	-	0.64	-
θ_Z	0.89	-	0.80	-	0.11	-	0.40	-
θ_M	0.11	-	0.12	-	0	-	0.06	-
θ_H	0	-	0.08	-	0.89	-	0.54	-
ε_Z	1	-	1	-	1	-	1	-
ε_M	0.56	-	0.83	-	-	-	0.13	-
ε_H	-	-	0.44	-	0.51	-	0.35	

Table 24: Network 2. Peripheral vs Central Players. Final Estimation

	Periphery LL=-1328.51				Centre LL=-1058.28			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.40	0.52	0.08	0	0.09	0.70	0.21	0
θ_Z	0.18	0.98	0.01	0.06	0.10	0.99	0.03	0
θ_M	0	0.01	0	0.06	0	0.01	0	0
θ_H	0.82	0	0.98	0.89	0.90	0	0.97	1
ε_Z	1	1	1	0.38	1	1	1	-
ε_M	-	1	-	0.89	-	1	-	-
ε_H	0.44	-	0.60	0.87	0.31	-	0.39	0.28

Table 25: Network 3. Peripheral vs Central Players. First Estimation.

	Periphery LL=-926.95				Centre LL=-2538.39			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.19	-	0.23	0.58	0.34	-	0.01	0.65
θ_Z	0.16	-	0.14	0.97	1	-	0.63	0.99
θ_M	0	-	0	0.03	0	-	0.29	0.01
θ_H	0.84	-	0.85	0	0	-	0.09	0
ε_Z	1	-	1	1	1	-	0.33	1
ε_M	-	-	-	0.67	-	-	0.52	1
ε_H	0.41	-	0.64	-	-	-	0.07	-

Table 26: Network 3. Peripheral vs Central Players. Second Estimation.

	Periphery LL=-926.95				Centre LL=-426.44			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.57	-	0.43	-	-	-	0.26	0.74
θ_Z	0.83	-	0.77	-	-	-	0.09	0.98
θ_M	0.17	-	0.10	-	-	-	0	0.01
θ_H	0	-	0.13	-	-	-	0.91	0
ε_Z	1	-	1	-	-	-	1	1
ε_M	0.61	-	0.82	-	-	-	-	0.45
ε_H	-	-	0.69	-	-	-	0.57	-

Table 27: Network 3. Peripheral vs Central Players. Final Estimation

C.5 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 28: Variants of belief-based learning

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

Note that your neighbours 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 neighbours you have as well as their experimental label, but not who they really are. This also means, of course, that your neighbours will not know your real label.

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

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 neighbours.
2. In the second stage you can request information about your neighbours, your neighbours' neighbours 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 neighbour's actions and payoffs in light grey. The table is read as follows (dark payoffs):

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

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

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

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

- the experimental label of your neighbours' neighbours (2nd order neighbours)
- the experimental label of your neighbours' neighbours' neighbours (3rd order)
- the experimental label of your neighbours' neighbour's neighbours' neighbours (4th order neighbours)

Note that who is a neighbour 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 neighbours' neighbours' you first need to request information about your neighbours 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 neighbours
- the actions chosen by your neighbours' neighbours
- the actions chosen by your neighbours' neighbours' neighbours
- the actions chosen by your neighbours' neighbour's neighbours' neighbours
- the payoffs obtained by your neighbours
- the payoffs obtained by your neighbours' neighbours
- the payoffs obtained by your neighbours' neighbours' neighbours
- the payoffs obtained by your neighbours' neighbour's neighbours' neighbours
- your own payoffs

Obviously, in order to receive information about your neighbours (or neighbours' neighbours') actions or payoffs you first need to request information about the experimental label of your neighbours (neighbours' neighbours) 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 neighbour. She chooses action B and you action D. Which gross payoff will you get in this round?
2. Assume you have three neighbours 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 neighbours change in every round of the game.
4. True or False: My neighbours face the same payoff table as I do.
5. True or False: My neighbours are those sitting in the cubicles to my left and right.