

Learning in Network Games*

Jaromír Kovářik^{†‡}

University of the Basque Country
and CERGE-EI

Friederike Mengel[§]

University of Essex

José Gabriel Romero[¶]

Universidad de Santiago

April 4, 2017

Abstract

We report the findings of experiments designed to study how people learn 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 these data to estimate learning types using finite mixture models. Monitoring information requests turns out to be crucial, as estimates based on choices alone show substantial biases. We also find that learning depends on network position. Participants in more complex environments (with more network neighbours) tend to resort to simpler rules compared to those with only one network neighbour.

JEL Classification: C72, C90, C91, D85.

Keywords: Experiments, Game Theory, Heterogeneity, Learning, Finite Mixture Models, Networks.

*We wish to thank Clayton Featherstone, Guillaume Frechette, Sanjeev Goyal, Nagore Iriberry, Matthew Jackson, Muriel Niederle, Alicia Perez-Alonso, Karl Schmedders, Aljaz Ule, Marco van der Leij, Nathaniel Wilcox, three anonymous Reviewers and audiences in Cambridge, Chapman University, Navarra, Stanford, 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), Jaromír Kovářik the Basque Government (IT-783-13) and the Spanish Ministry of Science and Innovation and European Fund Regional Development (ECO 2015-64467-R, ECO 2015-66027-P) and GACR (14-220445), and Gabriel Romero Fondecyt Chile (Grant 11121155) 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).

[‡]CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických veznu 7, 111 21 Prague, Czech Republic.

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

[¶]Departamento de Economía, Universidad de Santiago, 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. As examples, 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 the Prisoner’s dilemma, some forms of aspiration-based learning can lead to cooperation (Karandikar et al, 1998), while imitation and belief learning models 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. In the experiment participants play a network game over multiple rounds. We keep track of which information participants request between rounds. We then combine our knowledge of information requests with observed action choices to estimate a distribution of learning types using mixture models. 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 information is needed to make this choice.

Network games offer new opportunities to identify different learning rules. In networks (compared to random matching or fixed pairwise matching protocols) it is more often possible to distinguish learning models via information requests. As an example, consider myopic best response and forward-looking learning. Under random matching an agent needs to know the distribution of play in the previous period irrespective of whether she is myopic or forward-looking. In a network, though, a myopic best responder needs to know *only* the past behaviour of her first-order neighbours (who she interacts with), while a forward-looking learner may need to know the behaviour of her second-order neighbours to be able to predict what her first-order neighbours will choose in the following period.¹ An additional advantage of using networks is that doing so 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 also ask whether an agent’s position *within* a network affects the way she learns. Hence our study allows us to address two issues that previous studies have found difficult to address: identifying learning rules based on information requests and studying how stable learning is across differently structured social interactions.

In our main treatments, participants interacted in a 4×4 Anti-Coordination Game. Anti-coordination appears in many important economic situations such as congestion, pollution, oligopolistic (quantity) competition, immunization, provision of public goods, or whenever there are gains from differentiation. Compared to pure Coordination games, Anti-Coordination games have the advantage that different learning rules predict different choices more often (making identification easier) and, compared to e.g. conflict games, they have the advantage that standard learning models do not lead to limit cycles. Since Nash equilibria are attractors of many of the learning rules we will consider, these rules can only be distinguished using out-of equilibrium choices. With 4×4 games we hoped that, even if there is convergence to Nash equilibrium, that such convergence would not be immediate. Slow enough convergence is necessary to identify learning rules using out-of-equilibrium choices (Knoepfle et al., 2009).

Our analysis considers different learning types, where a subject’s type first determines her information requests, possibly with error, and her type and information requests jointly determine her

¹Which information she needs exactly will 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.

decision, again possibly with error. Learning types are drawn from a common prior distribution that we estimate using mixture models. We consider learning models from three prominent classes as possible descriptions of participants' behaviour: reinforcement learning, imitation as well as belief-based models, in particular myopic best response, different variants of fictitious play and forward-looking learning. These rules differ widely in their degree of sophistication. While forward-looking agents reason strategically, reinforcement learners do not even need to know that they are playing a game at all.²

The experiment consists of nine main treatments. Three treatments have endogenous information requests for three different network topologies. Three intermediate treatments on the same networks have endogenous information requests about actions and payoff, but *not* about the network structure. Those treatments were added to address potential endogeneity concerns with respect to network position in the fully endogenous treatments. We also conducted three treatments with the same networks but *without* information requests. In these full information treatments participants were given all the information that can be requested in the other treatments by default. We use these 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 pairwise differences between the distribution of action choices across these three treatment variations.³

We now briefly summarize our main results. First, we find that monitoring information requests is crucial. We compare our main results with simpler estimations based on action choices alone (i.e. ignoring information requests) and detect substantial biases in these estimates. Estimations based solely on observed action choices lead us to accept certain learning rules that participants 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 behaviour only, this problem is likely to be more severe in smaller 2×2 games often studied in experiments.

Second, we find that network position affects how people learn. Participants with only one network neighbor resort more often to myopic best response learning compared to others with more neighbors who tend to resort more often to reinforcement learning. Those results suggest that people adapt the learning heuristics they employ to the complexity of the environment. In particular, people facing more complex environments (such as having more network neighbors) seem to resort to simpler learning rules.

In terms of the specific rules participants use, there is substantial heterogeneity. However, most of our participants' decisions are best described by either reinforcement learning or myopic best responses. There is very little evidence of forward-looking behaviour and almost no evidence of imitation. This is true for all the networks we consider.

Since almost all our participants can be described by either reinforcement learning or (myopic) belief-based rules, our results support the assumptions of experience-weighted attraction learning (EWA, Camerer and Ho, 1998, 1999; 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 behaviour at the aggregate level, at the individual level fewer than 15% of our participants persistently request information consistent with *both* reinforcement learning

²All learning rules are described in detail in Section 4.

³In addition to these nine treatments, we also conducted three treatments using a Coordination Game and find evidence for the same rules as in the Anti-Coordination Game. We also conduct one treatment where participants play the Anti-Coordination Game bilaterally as a benchmark.

and belief-based learning rules.

In the following we relate our paper to the experimental literature on learning and also highlight some methodological differences between the existing literature and our approach. 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 behaviour of participants (see e.g. Erev and Roth (1998), Mookherjee and Sopher (1997), Apesteguia et al. (2007), Kirchkamp and Nagel (2007) or Feltovich (2000), among others). While the representative approach is simple and allows for parsimonious rules, one downside of this 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, Ho et al., 2008 or Frechette, 2009). 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.⁴ Our approach is conceptually different from this literature. Instead of trying to find one rule that describes the average or median behaviour of a population, we try to understand the distribution of learning rules.

One way to do so is to estimate learning models individually for each subject (Cheung and Friedman, 1997; Ho et al., 2008). This approach leads to consistent estimates if the experiment involves “sufficient” time periods in which participants learn, i.e. before the learning process has converged.⁵ If both cross-sectional and longitudinal variation is exploited, fewer periods of non-convergence are needed to avoid small sample biases. Such studies have been conducted by Camerer and Ho (1998, 1999), who assume that agents can be categorized in two subpopulations with different parameter values and Camerer et al. (2002), who estimate a mixture of standard and sophisticated EWA learning in the population. While the latter also estimate mixture models, the class of learning rules they consider is limited to EWA with its component rules reinforcement learning and fictitious play.⁶

There are two main differences between these studies and our research. First we keep track of which information participants request between rounds. This allows us to base our estimations on more than just choice data, making it easier to detect the underlying data generating process (Salmon, 2001). It also allows us to exploit the identification possibilities arising in network games if information requests are monitored. Second, our design allows for any kind of learning rule or heuristic *ex ante* and we restrict attention to certain rules only at the estimation stage. Since most other studies restrict information feedback given to participants, they effectively rule out some types of learning *ex ante*. If e.g. no information about payoffs of other participants is provided, then payoff-based imitation learning is impossible.

Our paper is methodologically closely related to El-Gamal and Grether (1995), Costa-Gomes et al. (2001; CCB, henceforth) and Knoepfle et al (2009). CCB also monitor information look-ups (using the software MouseLab) to study procedural models of decision making, where a participant’s type reflects their degree of sophistication. The main difference to our study is that they do not

⁴Imitation learning has been shown to explain behaviour well in e.g. Cournot games. See Apesteguia et al. (2007) as well as the literature cited therein. Reinforcement learning has done well in Erev and Roth (1998) and best response learning in Mookherjee and Sopher (1997). Forward-looking behaviour matches experimental data well in Mengel (2014).

⁵“Sufficient” can often mean practically infeasible in a typical experiment. Cabrales and García-Fontes (2000, Footnote 17) report that the precision of estimates starts to be “reasonable” after observing around 500 periods of play.

⁶Gill and Prowse (2015) estimate mixture models to study how cognitive and emotional skills affect level k learning.

study learning, but static decision rules instead. Knoepfle et al. (2009) study learning in different 4×4 normal form games using eye-tracking software. They find that, while eye-tracking seems to suggest that participants are quite sophisticated (similar to our forward looking types), their choices are more consistent with adaptive learning models. One of their conclusions is that, due to the large degree of noise in eye-tracking data, fundamental changes in the observational paradigm could help to differentiate genuine information lookups from noise fixations. Our paper provides a methodological contribution in this regard.

While there is a considerable amount of experimental research aimed at understanding how people learn in games, there is relatively little research on whether the same rules are used in different contexts or whether people adapt the heuristics they use in comparable environments of different complexity. Some studies ask whether learning differs across games. Camerer and Ho (1999), for example, estimate their EWA learning model on different classes of games (unique mixed equilibrium, coordination and dominance solvable). They find that EWA fits better than the comparison models across all classes, but there are some differences across games in estimated EWA parameters. Knoepfle et al. (2009) study learning in different 4×4 normal form games and find largely consistent patterns across games. In this paper, we also find evidence for broad consistency across games in our Anti-Coordination as well as three additional Coordination treatments, in that reinforcement learning and myopic best reply learning are the most important rules in both games considered. Grimm and Mengel (2012) find that learning is affected by the complexity of the environment (number of games), but they do not compare different learning models. To the best of our knowledge our paper is the first to study how the complexity of social interactions (measured by the number of neighbours in the network) affects learning in a given game.

The paper proceeds as follows. Section 2 describes in detail the experimental design. Section 3 gives an overview of behaviour using simple descriptive statistics. Section 4 introduces the learning models and the econometric framework. Section 5 contains the main results. Section 6 presents additional results and robustness checks and Section 7 concludes. Additional analyses, figures and tables as well as the experimental Instructions can be found in an Online Appendix.

2 Experimental Design

In this section we describe our design and provide details about the underlying two-player game, the networks and the information structure.

2.1 The Game

In most of our treatments participants repeatedly played the symmetric two player game depicted in Figure 1. We chose a 4×4 rather than a 2×2 game, because (i) we hoped that this would generate sufficiently slow convergence to equilibrium to be able to analyze learning in a meaningful way and since (ii) a larger game makes it easier to identify a larger number of different learning rules from observing agents' choices only. Hence, by choosing such a game we hoped to give good chances to estimations based on action choices alone.

Within each session the networks were fixed, which means that each participant played with the same first-order 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 period are given by the average payoff obtained in all the (bilateral) games against

	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

Anti-Coordination Game.

Figure 1: The Bilateral Anti-Coordination games used in the main treatments with pure strategy Nash equilibria (A,B); (B,A); (C,D) and (D,C).

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. Our main focus is on an Anti-Coordination game, since, as discussed above, in Anti-Coordination games different learning rules predict different choices more often compared to Coordination games (making identification easier). In Section 6.5 we discuss results of treatments based on a Coordination Game.

Treatments differed along two dimensions: network topology and information. 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), *M* (interMEDIATE) and *F* (Full information). In Subsection 2.2 we present our three network topologies and in Subsection 2.3 we explain the information conditions.

2.2 Network Topology

As we have argued in the Introduction, network games allow us to identify learning rules more easily compared to e.g. random matching and fixed pairwise matching protocols. Additional advantages of using networks (compared to pairwise matching protocols) include the fact that participants cannot trade-off different pieces of information as easily. For instance, knowing one's own action and the action of the opponent fully reveals the latter's payoff and vice versa in pairwise matching, but generally not in a network. There are also independent reasons to study learning in networks, as arguably, most real-life interactions take place via social networks. In our design we systematically manipulate the network topology.

To select network topologies we focus on one particular property of networks, namely the variance in degree. In networks with a low variance of the degree distribution, players tend to have a similar amount of neighbours, while in networks with a high variance in degree there will be some players with many network neighbours and others with few. We are interested in whether and how learning differs across these two types of players (few and many neighbours) and networks (low and high variance). The most symmetric situation we study is the circle network with zero variance in the degree distribution. Starting from the circle, we then increase the variance in degree (keeping some other network characteristics constant; see Table 1), thereby creating more asymmetric situations. Figures 2 – 4 show the three network architectures used in the experiment and Table 1 summarizes the most standard network characteristics of these networks.⁷

⁷Degree measures the number of first-order neighbours of a node and the clustering coefficient the share of a node's first-order neighbours who are neighbours themselves. The characteristic path length is the average length of shortest paths between any two nodes in the network and the betweenness centrality of a node measures the share of shortest

	Network 1	Network 2	Network 3
Number of players	8	8	8
Number of links	8	8	7
Average degree	2	2	1.75
Variance degree	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 1: Network Characteristics.

We used networks of 8 players, because in smaller networks identification of learning rules is harder. For example, in a circle of 3 players the sets of first-order neighbours and second-order neighbours coincide. In a circle of 4 (or 5) players the same is true for the sets of first- (second-) and third-order neighbours etc. In order to distinguish e.g. myopic best response learning from forward looking learning in terms of information requests these sets of neighbours should not overlap too much as we indicated above and will make clearer in Section 4. While many real-life networks will be even larger than 8 players, choosing larger networks in our experimental setting is likely to make the environment too complex for many participants. The trade-off between these two forces motivated us to choose networks of 8 players.

Network Equilibria 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. In the following, whenever we refer to equilibria we will refer to such network equilibria. Essentially there are two types of equilibria. In some all neighbours choose best responses. This means that neighbours alternate either between actions A/B or between actions C/D.⁸ But there are also network equilibria where agents only choose a best response to all neighbours *on average* but not to each neighbour individually. A table describing all strict Nash equilibria in the three networks can be found in Appendix A. Coordinating a network of 8 players on any one of several possible equilibria (between 9 – 12 depending on the network) is possible, but not obvious. We hoped to see mis-coordination in early periods and learning and convergence to equilibrium afterwards. The data show, that indeed, choices converge to a one-shot network equilibrium in several networks (see Section 3). It is important, though, to note that whether or not participants do converge to a Nash equilibrium does not matter for us *per se*, as long as they do learn. In fact, as highlighted above, observing some out-of equilibrium choices helps identifying learning rules.

paths between any two nodes that pass through the node in question. For formal definitions of these network properties see the textbook by Vega Redondo (2007).

⁸For example in Network 1 (see Figure 2) one network equilibrium is that players 1,...,8 choose actions $(a_1, \dots, a_8) = (A, B, A, B, A, B, A, B)$. 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).

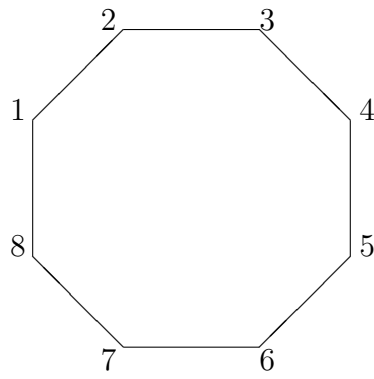


Figure 2: Treatments N-1, M-1 and F-1

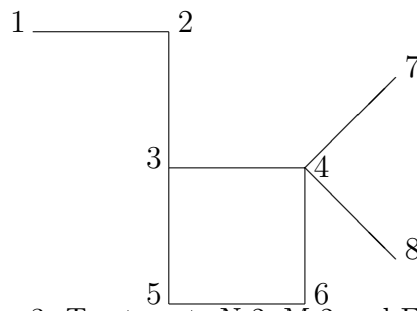


Figure 3: Treatments N-2, M-2 and F-2

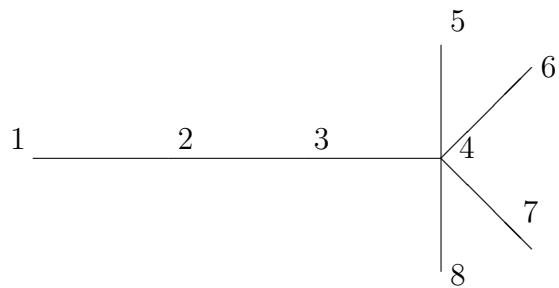


Figure 4: Treatments N-3, M-3 and F-3

2.3 Information

Our second treatment dimension varies how information about the network and histories of play is provided. We consider three variations labelled N (eNdoGenous), M (interMEdiate) and F (Full)

N-treatments In the *endogenous information* treatments ($N-1$, $N-2$ and $N-3$), we did not give our participants any information neither on the network nor on histories by default. Instead, at the beginning of each period 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. More precisely, if a participant requested the network position of her first-order neighbours she was shown how many neighbours she has and their experimental label (which is a number between 1 and 8; see Figures 2-4). With second-order neighbours, she was shown their experimental label as well as the links between the first- and second-order neighbours etc. Regarding actions and payoffs, participants were 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 participants held at any time of the experiment.

We placed two natural restrictions on information requests. First, participants were only allowed to ask for the actions and/or payoffs of neighbours 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. 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.⁹

M-treatments Data generated by our N -treatments will allow consistent estimation of learning rules, as long as the information collected does not affect a participant’s learning type. However, we have conjectured that the complexity of social interactions, measured by the number of network neighbours, may affect learning. If this was indeed the case, then learning type would be endogenous to the information collected about the network structure, leading to model misspecification.¹⁰ As one way to address this issue, we conducted the M -treatments ($M-1$, $M-2$ and $M-3$). They only differ from the N -treatments in that the network structure was known since the beginning of the experiment. Afterwards, participants could request information on past action choices and payoffs, as in the N -treatments. While the M -treatments avoid this possible endogeneity problem, the N -

⁹An 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 participants’ sequences are very noisy and highly heterogeneous” (p. 1209). Knoepfle et al (2009) who use eye-tracking software report similar issues.

¹⁰Since 90% of participants request information about the network structure by period 5 (see Section 3) in the N -treatments (and this information is permanently displayed once requested), this should not be too much of a problem. Still, the M -treatments described here can provide evidence on how important this issue is for the N -treatment estimates. We thank a Reviewer for suggesting these treatments.

treatments allow for better separation of learning rules. We will always show the results of both sets of treatments in parallel and discuss differences where they arise.

F-treatments Finally, to see whether information requests per se affect action choices (e.g. because participants might not request “enough” information due to the costs) we conducted three 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 of each period to all participants.

	Network 1	Network 2	Network 3	Bilateral
ENdogenous Information (N)	40 (800; 5)	56 (1120; 7)	40 (800; 5)	
InterMediate Information (M)	40 (800; 5)	32 (640; 4)	32 (640; 4)	
Full Information (F)	24 (480; 3)	24 (480; 3)	24 (480; 3)	26 (520, 13)
Total Anti-Coordination	104 (2080; 13)	112 (2240; 14)	96 (1920; 12)	26 (520, 13)
Coordination (M-Co)	32 (640; 4)	32 (640; 4)	40 (800; 5)	

Table 2: Treatments and Number of participants (Number of Observations; Networks/Pairs).

Four more treatments were run. In a bilateral treatment participants were matched in pairs to play the game for 20 rounds (see column “Bilateral” in Table 2). Further three treatments were conducted with a Coordination game (in the M -condition) to see how much learning changes across game forms. Results on the Coordination games are discussed in Section 6.5. Table 2 summarizes the treatment structure of the experiment.¹¹

All elements of the design were clearly explained in the Instructions, which can be found in the Online Supplementary Material posted with the replication files. After finishing the Instructions our participants had to answer several control questions regarding the understanding of the game, network interactions, information requests, and how payoffs are computed. There was no time constraint, but participants were not allowed to proceed without correctly answering all these questions. Experiments were conducted at the BEE-Lab at Maastricht University (N - and F -treatments) and at the Bilbao LABEAN at the University of the Basque Country (M -treatments, bilateral treatment, coordination treatments) using the software Z-tree (Fischbacher, 2007). A total of 442 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.

3 Action Choices and Information Requests

In this section, we provide a brief overview of action choices and information requests. Since our main interest is in estimating learning types, we keep this section short. Appendix A as well as the Online Supplementary Material contain many additional tables and details.

3.1 Action Choices and Payoffs

We analyze three aspects of choice behaviour, all related to the question of learning: (i) the frequency of switches, (ii) the evolution of payoffs over time and (iii) convergence to Nash equilibrium.

¹¹The table does not contain the treatment $N - 2$ without costs mentioned above. We will not discuss this treatment any further, but results are available upon request. Other than the treatments reported we did not conduct any other treatments or sessions and we did not run any pilot studies. The number of observations reported in the table differs across treatments, because of different “show-up rates” in the different treatments.

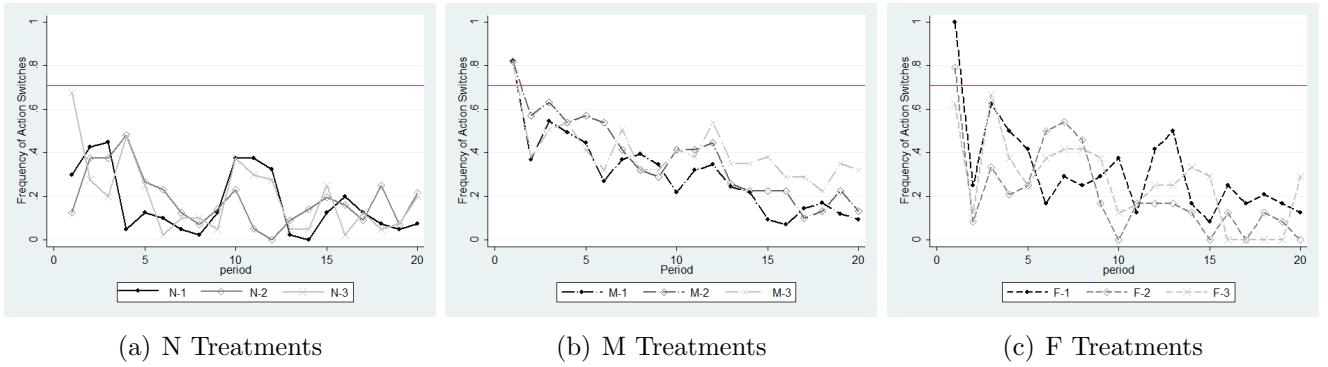


Figure 5: The frequency with which players change their action across rounds in different treatments. The horizontal reference line indicates the mean frequency of switches in the bilateral treatment.

Switching and Payoffs. Since it is only meaningful to analyze learning types if there is learning, we report some evidence concerning the stability of action choices. Figure 5 illustrates that players change their action less and less often over time, indicating that some convergence took place. Across periods switching decreases by 1.4-2.3 percentage points on average, depending on treatment. Switching is also substantially lower compared to the bilateral (pairwise matching) treatment. Participants also learned in the sense that they were able to increase their mean payoffs over time. Figure 6 shows that mean gross profits (i.e. game payoffs without taking into account costs of information requests) are increasing over time in all treatments. Payoffs range between two benchmarks: (i) the mean payoff in bilateral Nash equilibria (C,D) and (D,C) which is 35 and (ii) the payoff implied by random play of C and D, which is 25.

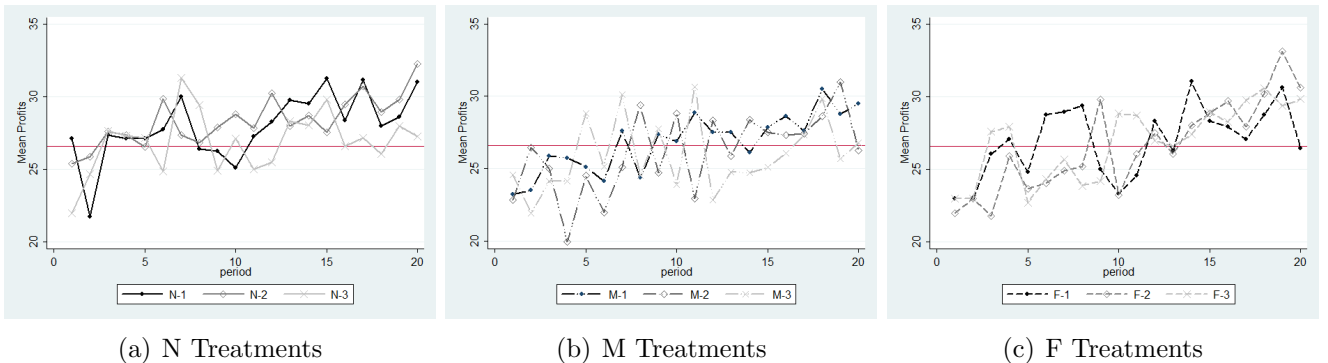


Figure 6: Mean *gross* payoffs across rounds in different treatments. The horizontal reference line indicates the mean payoffs in the bilateral treatment.

Nash equilibria. Players do not only switch less often and obtain higher payoffs over time, they also are more likely to play a Nash equilibrium of the network game over time. The entire network converges to an equilibrium between 0% – 46% of the time across the last 5 periods of play depending on treatment.¹² Participants never reach a Nash equilibrium in Network 3 and most often in Network 2. There seems to be no systematic effect of information conditions, with most coordination being observed in the *F*-treatments and least in the *M*-treatments (Tables 9 and 10 in Appendix A). In line with previous evidence on the conflict between risk and payoff dominance (e.g. Camerer, 2003), action choices always converge towards the “risk-dominant” actions *C* and *D* in our experiment.

¹²To help interpret these numbers, note that if players did choose actions uniformly at random the probability to coordinate on a Nash equilibrium would be 0.00018 for Network 1 and 0.00015 for Networks 2 and 3.

This also seems to be a result of learning, as the number of participants playing these actions is lower in the first half of the experiment across all three networks (see Table 11 in Appendix A). Full, InterMediate and Endogeneous Information treatments are statistically no different in terms of the distribution of action choices across the 20 periods (two sided ranksum test, $p > 0.2$).

Overall the evidence collected suggests that participants did learn in the experiment. Interestingly, having full information (as in the F - treatments) does not seem to improve payoffs or lead to less switching compared to the N - and M -treatments. This suggests that participants do request the information they need for learning in the M - and N -treatments and that information requests do not per se affect participants' behaviour.

3.2 Information Requests

This subsection focuses on information requests. We focus on simple averages in this subsection. Appendix C contains several figures showing the evolution of information requests over time.

Network Structure. In the first period 77.5%, 76.8% and 72.5% of participants in $N - 1$, $N - 2$ and $N - 3$, respectively, requested information about the experimental label of their first-order neighbours. Roughly 90% of individuals end up requesting this information by the fifth period of the experiment. Around 45% of participants request to know the network structure up to their second-order neighbours. Only 12.5%, 23.2% and 12.5%, respectively, request information about the entire social network. Remember that information about the network structure - once requested - was permanently displayed.

Payoffs. Around 50% of participants request their own payoffs in the N -treatments and around 30% do so in the M -treatments. Only between 8 – 12% of the time (depending on treatment) is information about the payoffs of first-order neighbours requested across all N - and M -treatments. For higher order neighbours these percentages are all below 2%.

Actions. Around 50% of participants request information about past actions of their first order neighbours in the N -treatments and around 40% in the M -treatments. Information requests about actions decline over time, which is consistent with the decline in switching behaviour observed in Figure 5 and could be attributed to convergence. Despite the strategic effect of second-order neighbours' action choices on the choices of direct opponents, the interest in their behaviour is relatively small (requested $\approx 10 - 15\%$ of the time).

4 Framework and Methods

This section discusses our selection of learning models and sets out basic issues in identifying learning rules from our data (i.e. from 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 Appendix B.1.

1. Under *Reinforcement Learning (RL)* participants in each period randomize between actions with probabilities that are proportional to the sum of past payoffs they obtained with these actions (Roth and Erev, 1995; Erev and Roth, 1998; Sutton and Barto, 1998; Skyrms and Pemantle, 2000; Hopkins, 2002).¹³

¹³In our estimations we will assume that a participant perfectly consistent with RL chooses the most preferred action with probability one. This approximates some exponential choice rules used in the literature, but is not the case with e.g. the linearly proportional rule.

2. Under *Payoff-based Imitation (PBI)* participants choose the action with the highest average payoff in the previous period in their first-order neighbourhood including themselves (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).¹⁴
3. Under *Myopic Best Responses (MBR)* players choose a myopic best-response to the distribution of 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)* players assume that their first-order neighbours are myopic best responders and best-respond to the anticipated distribution of choices in the following period (Blume, 2004; Fujiwara-Greve and Krabbe-Nielsen, 1999; Selten, 1991; Mengel, 2014).¹⁵

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, aspiration-based reinforcement learning and several variants of payoff-based imitation (including imitation of different sets of players, e.g. higher-order neighbours). These rules only differ from the above rules in few instances of predicted action choices and information requests. The four rules singled out above are each representative of a larger class of learning models. Including all possible variants would (a) over-specify the model considerably and (b) lead to many instances of non-identifiability (where two, possibly quite similar, rules prescribe both the same action choice and information requests).

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 behaviour by looking at how well its component rules perform. The reader may also wonder why we did not include level- k learning rules. 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 how level- k types should update their beliefs about the distribution of levels k in the population upon receiving new information.¹⁶ There is a connection, though, between our myopic best response type and level-1 and the forward looking type and level-2 in that the latter reasons about how their opponents learn (choose), while the former takes opponents' choices as given.

4.1 Identifying Learning Rules from Data

It is only possible to identify learning rules if different rules imply different choices and/or information requests in the experiment. The average number of periods (out of 19) in which two different learning types predict different action choices for a participant given the history of play ranges between 7-15 periods for our rules (Table 19 in Appendix D). The fact that the learning rules considered entail different predictions quite often is due to our design involving the 4×4 Anti-Coordination game

¹⁴Some 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.

¹⁵Tang (2001) experimentally tested the model of anticipatory learning by Selten (1991).

¹⁶One proposal in this vein has been made by Knoepfle et al (2009) who model a sophisticated learning type (similar to our forward looking type) which exhibits a recursive structure similar to level- k models starting from Cournot best response as the anchoring C0 (level-0) type. Using eye-tracking technology, they find information look up patterns consistent with such a learning type, though choices are better explained by adaptive models. Gill and Prowse (2015) estimate a mixture model based on level- k types to analyze how cognitive ability and character skills affect equilibrium play.

and should give good chances to estimations based on action choices alone. We will see below that, despite this fact, estimates are substantially biased if only action choices are considered.

Apart from choices we also observe participants' information requests. Figure 7 provides an example of how different learning rules imply different information requests for player 7 in Network 1. As a reinforcement learner she does not need to know anything about the network or the choices of others. In fact she does not even have to know the payoff matrix. She only has to know the payoffs she obtained in the past with each of the different actions. A payoff-based imitator should have information about the choices of her first-order neighbours and the payoffs they obtained in the previous period.¹⁷ Under myopic best responses she would need to know past choices of her first order neighbours and as a forward looking learner she would also need to know her second order neighbours and their choices in the previous period. Table 20 in Appendix D summarizes the information requirements for each rule.

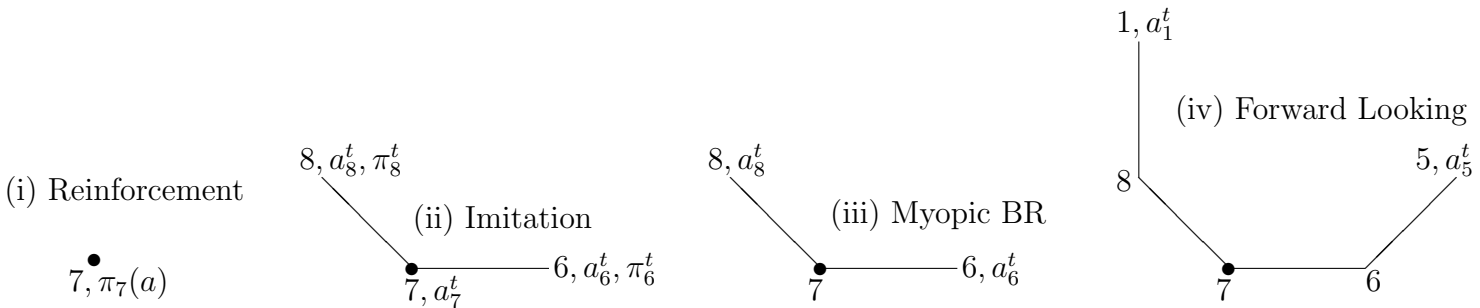


Figure 7: Player 7 in Network 1. Information required for Rules (i) RL, (ii) PBI, (iii) MBR and (iv) FL. a_i^t denotes the action taken by player i at time t . π_i^t the payoffs obtained by player i at time t and $\pi_i(a)$ the vector of average payoffs obtained by player i with each of the four actions.

If participants always requested at least the minimal information needed for a learning type, then all rules could be identified in all of the 20 periods in the N -treatments. 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 in the N -treatments 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). Because in the M -treatments participants do not request information about labels, identification is not as good. In particular reinforcement learners can infer their payoffs also from their neighbours' action choices without having to request their labels in addition. (The converse is not true, belief learners cannot - in general - identify choices of neighbours from their own payoffs). Since we can exploit correlations between information requests and implied choices, identification is still reasonably good even in the M -treatments.

¹⁷Note that we have not included a participant's own payoff in the minimal information set for PBI. The reason is that this information can be inferred from observing one's own and neighbour's action choices. The *minimal information set* we are looking for hence excluded this information. We have, however, run all estimations also with the larger set that requires a PBI type to request their own payoffs as well. Results do not change qualitatively and quantitative changes are minimal.

4.2 Econometric Framework

In this section we introduce the econometric framework. Our aim is to estimate population shares p_k of the different learning rules k discussed above. To do so we will estimate mixture models to try and find the distribution $p = (p_1, p_2, \dots, p_K)$ that can best explain our data, i.e. action choices and information requests. We start with the following assumption, which links information requests to learning rules:

Occurrence: In every period 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. On the other hand, unlike in eye-tracking studies, it is unlikely that occurrence is satisfied by chance, since participants had to pay for each piece of information they requested (see Section 2).

We then proceed as follows. For each subject i and learning type $k \in \{1, 2, \dots, K\}$, we compute the number of periods I_{kO}^i , in which subject i asked at least for the minimum information required to be consistent with learning rule k . Subscript O stands for ‘‘Compliance with Occurrence’’; $I_{kZ}^i = 19 - I_{kO}^i$ measures the number of periods subject i did not ask for the minimum information set corresponding to rule k , where subscript Z stands for ‘‘Zero Compliance with Occurrence’’. Let θ_{kj} denote the probability that a participant has compliance j with rule k in the experiment, where $j \in \{Z, O\}$ and $\theta_{kZ} + \theta_{kO} = 1$ for each k . Hence $\theta_{kZ} = 1 - \theta_{kO}$ is the probability that a participant of type k does *not* request the information required for k . We define $\theta_k = (\theta_{kO}, \theta_{kZ})$.¹⁸

We now turn to action choices. For a given subject in a given period, 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 action choices consistent with a given learning rule in a given period. A subject employing rule k normally makes decisions consistent with k , but in each period, given compliance j she makes an error with probability $\varepsilon_{kj} \in [0, 1]$. We assume that error rates are i.i.d across periods and participants. 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 a given action consistent with rule k (either by mistake or as a result of employing rule k) is

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

The probability to choose a given action that is inconsistent with rule k is $\frac{\varepsilon_{kj}}{4}$.¹⁹ We define $\varepsilon_k = (\varepsilon_{kZ}, \varepsilon_{kO})$. Under the assumptions made (low) compliance (‘‘errors in information requests’’) and action errors will be positively correlated. This is shown in Appendix B.2. In fact, for our main estimates we will even assume that in the case of zero compliance with rule k , participants can only be consistent with k by mistake, though dropping this assumption only leads to minimally changes in estimates. Conditional on j and k , however, errors are i.i.d across periods and participants.

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 I_{kj}^{ic} denote the number of periods in which subject i has compliance j with learning type k and c action choices are consistent with k . Hence $\sum_c I_{kj}^{ic} = I_{kj}^i, \forall i, k$ and j . x_{kj}^{ic} denotes the number of periods in which i takes one of c possible action choices consistent

¹⁸In Section 6.3 we increase the number of compliance categories.

¹⁹Note that $c\left(1 - \frac{4-c}{4}\varepsilon_{kj}\right)\frac{1}{c} + (4-c)\frac{\varepsilon_{kj}}{4} = 1$.

with k and has compliance j with k . Define $\sum_c x_{kj}^{ic} = x_{kj}^i$ for all i, k and j ; $x_k^i = (x_{kZ}^i, x_{kO}^i)$, $I_k^i = (I_{kZ}^i, I_{kO}^i)$; $I^i = (I_1^i, \dots, I_K^i)$, $x^i = (x_1^i, \dots, x_K^i)$; $\mathfrak{S} = (I^1, \dots, I^N)$ and $X = (x^1, \dots, x^N)$. As a result, the probability of observing sample (I_k^i, x_k^i) when participant i is of type k is

$$L_k^i(\varepsilon_k, \theta_k | I_k^i, x_k^i) = \prod_j \prod_c \theta_{kj}^{I_{kj}^{ic}} \left[\left(1 - \frac{4-c}{4} \varepsilon_{kj} \right) \frac{1}{c} \right]^{x_{kj}^{ic}} \left(\frac{\varepsilon_{kj}}{4} \right)^{I_{kj}^{ic} - x_{kj}^{ic}}, \quad (2)$$

and the log-likelihood function for the entire sample is

$$\ln LF(p, \varepsilon, \theta | \mathfrak{S}, X) = \sum_{i=1}^N \ln \left\{ \sum_{k=1}^K p_k \prod_j \prod_c \theta_{kj}^{I_{kj}^{ic}} \left[\left(1 - \frac{4-c}{4} \varepsilon_{kj} \right) \frac{1}{c} \right]^{x_{kj}^{ic}} \left(\frac{\varepsilon_{kj}}{4} \right)^{I_{kj}^{ic} - x_{kj}^{ic}} \right\}. \quad (3)$$

Under mild conditions satisfied by (3), the maximum likelihood method produces consistent estimators for finite mixture models (Leroux, 1992). Our aim is to find a mixture model $p = (p_1, p_2, \dots, p_K)$ that provides the best evidence in favor of our data set. With K learning types, we have $(4K - 1)$ free independent parameters: $(K - 1)$ independent probabilities p_k , K information request probabilities θ_{kj} , and $2K$ error rates ε_{kj} .

It is well known that testing for the number of components in finite mixture models “is an important but very difficult problem, which has not been completely resolved” (MacLachlan and Peel, 2000, p. 175). Standard information criteria for model selection, such as the likelihood ratio test, Aikaike or Bayesian Information Criteria, might not perform satisfactorily (Prasad et al., 2007, Cameron and Trivedi, 2005, 2010). In the literature there are two different approaches to deal with model selection (i.e. selection of components). Cameron and Trivedi (2005) propose to use the “natural” interpretation of the estimated parameters to select components, while MacLachlan and Peel (2000) argue that the true number of components generating the data is the smallest value of K such that the estimated likelihoods differ across components and all the p_k are non zero. In the following, we use a procedure that combines these two criteria. 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:

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

The left hand side of (4) is decreasing in the error rate, approaching 0 as $\varepsilon_{k,j}$ tends to 1. This means that choices consistent with type k are taken as evidence of learning rule k only if the estimated error rates suggest that those choices 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 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 θ_{kO} is high. 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_{kO} indicate that participants do not consult the minimum information corresponding to rule k very often and it would be hard to argue that learning rule k explains the behaviour of the participants clustered in component k . In other words, if θ_{kZ} is very high people classified as type k almost never consult the information corresponding to rule k and this is evidence that their learning behaviour was not actually generated by type k , irrespective of the estimated p_k .

With these considerations in mind, we will use the estimated values of θ_k as a tool for selecting the components of our finite mixture model. In particular we will (i) repeatedly estimate the multiple

learning rule model and (ii) eliminate after each repetition the rule l with the highest estimated θ_{lZ} whenever it is larger than an elimination threshold $\bar{\theta}_Z$. The process stops with the final estimation when either all remaining rules have an estimated $\theta_{lZ} < \bar{\theta}_Z$ or when only one rule remains. We would also have eliminated rules if estimated error rates did not increase as compliance decreases, but this did never happen in our data.

The elimination threshold $\bar{\theta}_Z$ can in principle be set to any level depending on when one starts to believe that a rule fails to explain behaviour. The rules we eliminate have an estimated $\theta_{lZ} > 0.9$, meaning that participants clustered in these rules only consult the information required for these rules with less than 10 percent probability.

Robustness Checks We also artificially altered the order of elimination of the types (for which $\theta_{lZ} > \bar{\theta}_Z$) and in all cases we converge to the same mixture composition as in the benchmark cases (see Online Supplementary Material). Hence, the results are robust to the order of elimination of learning types. They are also robust to alternative assumptions on the data generating process, to changing the number of compliance categories and relaxing the assumption of compliance with occurrence (Section 6). We now proceed to presenting our main results.

5 Main Results

5.1 Results based on Information Requests and Choices

This section contains our main results. We start by illustrating how our algorithm selects components using treatment $N-1$ as an example. Table 3 shows the estimated type frequencies p_k and parameters θ_{kZ} . After the first estimation with all rules we find that $\hat{\theta}_{PBI,Z} = 1$ (in bold in Table 3), meaning that participants classified as *PBI* do *not* consult the information required by this learning rule ever.²⁰ Therefore, our selection criterion suggests that there is no evidence that participants' choice behaviour was induced by *PBI* and we remove *PBI* from the estimation. In the second iteration we eliminate the forward-looking rule with $\hat{\theta}_{FWL,Z} = 0.99$. The algorithm stops with only two rules, *RL* and *MBR*, remaining. The selection algorithm selects the same learning rules in $N-2$ for all thresholds $\bar{\theta}_Z \in [0.48, 0.97]$ and in treatment $N-3$ only *MBR* survives for all $\bar{\theta}_Z \in [0.17, 0.97]$. We describe the results in more detail below.

How can it be that at the first step of estimations a rule that clearly does not describe behaviour well, such as *PBI*, obtains an estimated value of $p_{PBI} = 0.55$? Remember that the estimation procedure favours a rule k if its compliance with Occurrence is more concentrated on one particular value. Hence if participants' choices explain the variation in information requests poorly, this will lead to a high concentration on zero compliance (high θ_{kZ}) and will favor the estimated value of p_k . For this reason any estimated value of p_k can only be interpreted jointly with the vector θ_k (see also the discussion in Section 4.2).

There is information, though, that can be gained by studying the sequence of estimations (see Online Supplementary Material). In $N-3$, for example, our population is overall best described by *MBR*. But small percentages of decisions are also very accurately described by other rules that eventually get eliminated by the algorithm. For example 15% are very accurately described by reinforcement learning with $\hat{\theta}_{RL,O} = 0.85$ (first iteration). Hence, while our selection algorithm forces 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

²⁰The acronyms PBI, FL, RL and MBR stand for “payoff based imitation”, “forward looking”, “reinforcement learning” and “myopic best responses”, respectively (see Section 4).

<i>Parameters</i>	<i>Learning Types</i>			
	<i>RL</i>	<i>PBI</i>	<i>MBR</i>	<i>FL</i>
<i>1st iteration</i>				
p_k	0.25	0.55	0.20	0
θ_{kZ}	0.04	1	0.07	-
<i>2nd iteration</i>				
p_k	0.18		0.24	0.58
θ_{kZ}	0.09		0.07	0.99
<i>Final iteration</i>				
p_k	0.57		0.43	
θ_{kZ}	0.56		0.10	

Table 3: Estimation results after different iterations of the selection algorithm. Treatment $N - 1$.

rules are able to explain (a small) part of the data accurately and which rules can best account for the more noisy decisions.

Parameters	Treatments					
	N-1		N-2		N-3	
	RL	MBR	RL	MBR	MBR	FL
p_k	0.57***	0.43***	0.24*	0.76***	0.27***	0.73***
(s.e.)	(0.10)	(0.10)	(0.16)	(0.17)	(0.08)	(0.08)
θ_Z	0.56*	0.10	0.05	0.47**	0.16***	0.97***
(s.e.)	(0.30)	(0.26)	(0.27)	(0.20)	(0.05)	(0.04)
θ_O	0.44	0.9***	0.95***	0.53***	0.84***	0.03
(s.e.)	(0.30)	(0.26)	(0.27)	(0.20)	(0.05)	(0.04)
ε_Z	1***	1***	1***	1***	1***	1***
(s.e.)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
ε_O	0.51***	0.45***	0.48***	0.43***	0.64***	0.72***
(s.e.)	(0.09)	(0.11)	(0.08)	(0.07)	(0.08)	(0.13)
ln LL	-1325		-1861		-1203	
Individuals	40		56		40	
Observations	760		1064		760	

Table 4: Main Estimation Results using *both* information requests and action choices. **N-treatments**. Note: (***) significant at 1%, (**) at 5% and (*) at the 10% level. Standard errors computed using bootstrapping method with 500 replications following Efron and Tibshirani (1994). Note: The rule FL is eliminated in this stage in N-3 as $\theta_Z > 0.9$.

Table 4 reports the maximum likelihood estimates of learning type probabilities, p_k , compliance probabilities, θ_{kj} , compliance conditional error rates, ε_{kj} , and the corresponding standard errors (s.e.) in the selected models for the N -treatments. 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 ($\hat{\theta}_{RL,O} = 0.44$), while $\hat{\theta}_{MBR,O}$ even equals 90%. In both cases, estimated error rates increase as compliance decreases (i.e. the more frequently people classified as type k request information consistent with k , the more frequently their choices are consistent with rule k). These results suggest that the estimated type frequencies of RL and MBR are reliable. In $N - 2$, 24% and 76% of participants are best described by RL and MBR , respectively. The estimated θ 's and ε 's are also well behaved. For thresholds below 97% only MBR survives in $N - 3$. In the final estimation 73% of participants are classified as forward-looking learners, but they request information

Parameters	Treatments					
	M-1		M-2		M-3	
	RL	MBR	RL	MBR	RL	MBR
p_k	0.51***	0.49***	0.31**	0.69***	0.19**	0.81***
(s.e.)	(0.00)	(0.00)	(0.13)	(1.35)	(0.07)	(0.08)
θ_Z	0.20***	0.90***	0.22	0.68***	0.09	0.68***
(s.e.)	(0.040)	(0.05)	(0.13)	(0.07)	(0.11)	(0.05)
θ_O	0.80***	0.10**	0.78***	0.14*	0.91***	0.12**
(s.e.)	(0.04)	(0.05)	(0.13)	(0.07)	(0.11)	(0.05)
ε_Z	1***	1***	1***	1***	1***	1***
(s.e.)	(0.06)	(0.00)	(0.04)	(0.04)	(0.04)	(0.04)
ε_O	0.57***	0.48***	0.6***	0.49***	0.84***	0.63***
(s.e.)	(0.00)	(0.15)	(0.11)	(0.13)	(0.11)	(0.14)
ln LL	-1316		-1082		-1060	
Individuals	40		32		32	
Observations	800		640		640	

Table 5: Main Estimation Results using *both* information requests and action choices. **M-treatments.** Note: (***) significant at 1%, (**) at 5% and (*) at the 10% level. Standard errors computed using bootstrapping with 500 replications (Efron and Tibshirani, 1994).

consistent with this rule (i.e. their second-neighbors' action choices) only with probability 0.03.

Table 5 shows the analogous results for the M -treatments. In all three M -treatments, the population is described as composed by RL and MBR . In $M - 1$ about equally many participants are classified as RL (51 %) and MBR (49 %), even though the MBR estimate is very noisy. In $M - 2$ the percentage classified as MBR increases to 69% and it even reaches 81% in $M - 3$ (with the remainder being classified as RL). Across all treatments RL types are identified more precisely than MBR types, suggesting that many of the more noisy learners are classified as MBR . This could possibly be due to the slightly worse identification in the M - treatments discussed above. Some participants request their neighbours' action choices may infer their own payoffs and hence display choices less consistent with MBR .²¹ Overall the N - and M -estimates deliver a similar picture. Reinforcement learning and myopic best response learning are the two rules selected in Networks 1 and 2 with a higher share of belief learning in Network 2. In those two networks estimated shares are less than 7 percentage points apart across the two information conditions. There are seemingly more differences in Network 3. In $M - 3$ reinforcement learning receives a share of 19% while in $N - 3$ all participants are classified as belief learners. Remember, though, that in the initial step of elimination 15% of decisions were very accurately described by reinforcement learning even in $N - 3$ (Online Supplementary Material).

Since almost all our data can be described by either reinforcement learning 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 -

²¹Because of our assumption that a participant who does *not* request information consistent with k can only choose an action consistent with rule k by chance and not as a result of employing rule k , the estimated $\hat{\varepsilon}_{kZ} = 1$ for all networks and selected rules. If we drop this assumption, then the p_k estimates are the same in $N - 1$, $N - 3$ and $M - 3$, we find slightly higher frequencies of RL in $M - 1$ and $M - 2$ and more evidence of MBR in $N - 2$. In that case the estimated ε_{kZ} range between 0.51 and 1 for the selected rules.

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 behaviour at the aggregate level, at the individual fewer than 15% of our participants persistently request information consistent with both reinforcement learning and belief-based learning rules (see Figure 12 in Appendix C). There is also a connection between our findings and the idea of multi-dimensional iterative reasoning introduced by Arad and Rubinstein (2012). Our participants who are best described by MBR (as opposed to FL) are myopic (level 1) both in terms of assuming stationary behaviour of neighbours and in terms of only reasoning about first-order neighbours, but not higher order neighbours.

Some readers might wonder whether we are overestimating the frequency of RL , because participants might look up their own payoffs just because they want to know their payoffs and not because they use this information in their learning rule. We probably do, but only to a small extent. Note, first, that the estimation procedure identifies high correlations between information 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 find no evidence for RL in $N - 3$ indicates that this is a minor issue in our study.

The comparison of the three networks suggests that network topology affects how people learn. Both in the N - and M -treatments, we observe an increase of belief-based models at the expense of the simpler reinforcement learning as we move from Network 1 to Network 3. One possible reason for this pattern could lie in the fact that in Network 3 there are many (5) network positions with only one network neighbour, there are some (3) in Network 2, but none in Network 1. A conjecture we will evaluate in Section 5.3, below, is that players in simpler environments (i.e. with fewer network neighbours) rely on more sophisticated learning rules, while players in more complex environments tend to resort to simpler rules, such as reinforcement learning.

5.2 Results based only on Action Choices

In this section we will try to understand how much is gained by using the methodology outlined in the previous subsection compared to simpler estimations based on action choices alone. If results obtained via the latter set of estimations are “worse” than those obtained via our main estimations, then (at least in this context) collecting the additional information seems crucial and the advantage of the network approach would be highlighted. Hence, the objective is to test whether estimates are less accurate if data on information requests are ignored.

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 . Let $T_k^{i,c}$ be the number of periods in which subject i has c possible action choices consistent with rule k . And let $x_k^{i,c}$ measure the number of periods 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 then be written as

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

The log-likelihood function is:

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

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. The parameters of equation (6) are estimated using maximum likelihood methods as before. Now we have $2K - 1$ free independent parameters, $(K - 1)$ corresponding to frequency types p_k , and K corresponding to the error rates.

Table 6 reports the estimated frequencies and error rates. There is evidence in favor of all four learning types. Based on these results we could conclude that there is evidence of payoff-based imitation in $N - 3$ (23%), $M - 1$ (20%), $M - 2$ (53%) and $M - 3$ (50%) even if we have already seen above (Section 4) that action choices and information requests are inconsistent with *PBI*. We also obtain a significant share of *FL* (30% for $N - 1$ and 11% in $N - 2$) despite the fact that participants hardly ever requested information needed to identify the corresponding action choices. Consequently, it is very unlikely that these learning rules have generated the behaviour of participants in the experiment.

Parameters	<i>Treatment N-1</i>				<i>Treatment M-1</i>			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.21***	0.08	0.42***	0.30***	0.46***	0.20**	0.24***	0.10*
(s.e.)	(0.08)	(0.07)	(0.12)	(0.10)	(0.09)	(0.08)	(0.09)	(0.05)
ε_k	0.08	1***	0.58***	0.42***	0.35***	0.87***	0.57***	0.09
(s.e.)	(0.16)	(0.21)	(0.16)	(0.07)	(0.06)	(0.08)	(0.08)	(0.20)
ln LL		-760				-794		
Individuals		40				40		
Observations		760				760		
Parameters	<i>Treatment N-2</i>				<i>Treatment M-2</i>			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.49***	0.04	0.35***	0.11***	0.17*	0.53***	0.24***	0.06
(s.e.)	(0.08)	(0.04)	(0.08)	(0.06)	(0.10)	(0.16)	(0.09)	(0.13)
ε_k	0.26***	0.48**	0.47***	0.76***	0.33**	0.87***	0.41***	0.03
(s.e.)	(0.04)	(0.25)	(0.08)	(0.19)	(0.17)	(0.11)	(0.09)	(0.46)
ln LL		-1022				-709		
Individuals		56				32		
Observations		1064				608		
Parameters	<i>Treatment N-3</i>				<i>Treatment M-3</i>			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.51***	0.23**	0.21***	0.05	0.33**	0.50***	0.08	0.09
(s.e.)	(0.09)	(0.08)	(0.08)	(0.05)	(0.15)	(0.17)	(0.13)	(0.11)
ε_k	0.34***	0.70***	0.42***	0.28	0.48***	0.88***	0.27	0.15
(s.e.)	(0.06)	(0.21)	(0.07)	(0.26)	(0.15)	(0.22)	(0.31)	(0.32)
ln LL		-772				-718		
Individuals		40				32		
Observations		760				608		

Table 6: Estimation based solely on observed action choices. Note: (***) significant at 1% level; (**) at 5% level; (*) at 10% level. Standard errors computed using bootstrapping methods with 500 replications (Efron and Tibshirani, 1994).

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 those rules. The estimation procedure identifies high correlations between information requests 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 experimental design renders (ii) unlikely, and (i) is only very rarely the case in our experiment (Table

19 in Appendix D). Note also that situations such as (i) will likely affect estimations that disregard information requests even more.

We conducted analogous estimations for the full-information treatments (Table 24 in Appendix D). Note that in these treatments, since all information is available, we do not know which information participants used in their decisions. Any inference, hence, has to be made purely on the basis of observed action choices. The estimates show no systematic shift of learning toward any rule when comparing these tables. This provides further evidence that the “small” costs imposed on information requests did not distort the way participants learn (see also Section 4).

To sum up, we have seen that straightforward maximum likelihood estimation that disregards information requests ends up accepting learning rules for which participants would need information that we know they did not have. As our design (involving the 4×4 Anti-Coordination game) was chosen in order to give estimations using action choices alone good chances to detect learning strategies (see Section 4.1), one might expect these biases to be much more severe for smaller games or pure Coordination games, where identification based on choices alone is more difficult.

5.3 Estimates by Network Position

In order to understand whether learning is affected by network position, we estimate our model separately for different groups of network positions. In particular we ask whether the number of network neighbours affects how people learn. Since having more neighbours involves collecting and processing more pieces of information and hence higher cognitive costs, players with many neighbours might resort to rules that are less demanding in terms of information requirements.²² Since reinforcement learning requires storing and processing one piece of information (own payoff) irrespective of the number of neighbours, it is arguably less costly in terms of cognitive resources to resort to this rule in positions with many neighbours. The number of different pieces of information a decision maker needs to process under MBR learning, however, is linearly proportional to the number of neighbours (action choices for each neighbour).²³

To estimate the model separately for each position in the networks would lead to very small samples and likely to small-sample biases. We hence 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. To have balanced groups we omit Network 1, but it is included in Tables 22 and 23 in Appendix D without substantially affecting results. Group 1 (with one network neighbour) contains players 1,7, and 8 in Network 2 and 1,5,6,7, and 8 in Network 3, whereas Group 2 (multiple network neighbours) contains players 2,3,4,5, and 6 in $N - 2$ and 2,3, and 4 in Network 3. Table 7 reports the estimation results.

The estimates seem to support our conjecture. More complex situations (more network neighbours) trigger the use of less sophisticated rules. In the N -treatments (reported in Table 7) reinforcement learning gets attributed somewhat more and belief learning somewhat less than 50% in Group 1. In Group 2, by contrast, the shares attributed to reinforcement learning (65%) are almost double those of *MBR* (35%). The difference between Group 1 and Group 2 estimates is statistically highly significant ($z = 5.058, p < 0.001$).²⁴ In the M -treatments somewhat more participants are

²²In a different context Brock and Hommes (1997) show that acknowledging agent’s choice between simple and more complex learning or predictor rules can dramatically alter our understanding of equilibrium (or non-equilibrium) behaviour.

²³In particular agents with more than one neighbour need to consider multiple past/future action choices and take weighted averages to make their choice (under the more sophisticated rules *MBR* and *FL*) while agents with only one neighbour do not need to do this. Note, though, that only *one* request is needed to receive information about choices for all neighbours.

²⁴The bootstrapped estimates return a covariance between the estimated p_{RL} across the groups of < 0.0007 . We

classified as *MBR* in both groups. But again, the estimated population share of *RL* is higher in the more complex environment, with 41 % being classified as *RL* in Group 2 and only 23 % in Group 1 ($z = 22.220, p < 0.001$).

<i>N-treatments</i>								
Parameters	Group 1: one network neighbor				Group 2: multiple network neighbors			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.54***		0.46***		0.65***		0.35*	
(s.e.)	(0.10)		(0.10)		(0.18)		(0.18)	
θ_Z	0.87***		0.74**		0.61*		0.11	
(s.e.)	(0.26)		(0.33)		(0.37)		(0.21)	
θ_O	0.13		0.26		0.39		0.89***	
(s.e.)	(0.26)		(0.33)		(0.37)		(0.21)	
ε_Z	1***		1***		1***		1***	
(s.e.)	(0.04)		(0.04)		(0.04)		(0.04)	
ε_O	0.63***		0.7***		0.49***		0.35***	
(s.e.)	(0.12)		(0.11)		(0.16)		(0.07)	
ln LL			-1630				-1652	
Individuals			46				50	
Observations			874				950	
<i>M-treatments</i>								
Parameters	Group 1: one network neighbor				Group 2: multiple network neighbors			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.23***		0.77***		0.41***		0.59***	
(s.e.)	(0.09)		(0.09)		(0.117)		(0.117)	
θ_Z	0.21		0.86***		0.27**		0.95***	
(s.e.)	(0.149)		(0.053)		(0.118)		(0.072)	
θ_O	0.79***		0.14***		0.73***		0.05	
(s.e.)	(0.149)		(0.053)		(0.118)		(0.072)	
ε_Z	1***		1***		1***		1***	
(s.e.)	(0.045)		(0.045)		(0.065)		(0.045)	
ε_O	0.68***		0.64***		0.69***		0.49**	
(s.e.)	(0.155)		(0.115)		(0.091)		(0.203)	
ln LL			-1091				-1051	
Individuals			32				32	
Observations			640				640	

Table 7: Participants with one network neighbour (1,7,8 in M-2 and 1,5,6,7,8 in M-3) vs participants with multiple network neighbours (2,3,4,5,6 in M-2 and 2,3,4 in M-3). Note: (***) significant at 1% level, (**) at 5% level and (*) at 10% level. Standard errors computed via bootstrapping with 500 replications (Efron and Tibshirani, 1994).

To sum up, these results suggest that learning depends on network position. In particular, the complexity of the environment measured by the number of network neighbours seems to affect which learning rules participants resort to.

6 Additional Results

In this section we report several robustness checks. First, using simulated data we evaluate to what extent our econometric model is capable of identifying the learning rules present in the population. Second, we discuss specifications with alternative learning rules. Third, we study alternative assumptions on compliance with occurrence. Fourth, we discuss alternative assumptions on the data generating process. Last, we report results on our Coordination Game treatments. For reasons of space all robustness checks in Sections 6.1-6.3 are reported only for the *N*-treatments, but they all extend to the *M*-treatments.

hence treat them as independent and test the null-hypothesis of equality of coefficients across groups 1 and 2 using a z-statistic which accounts for degrees of freedom across the groups.

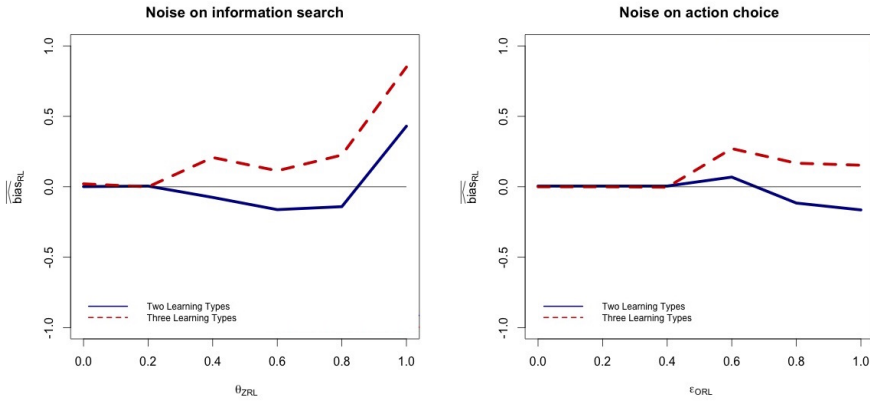


Figure 8: Difference between estimated parameters \hat{p}_{RL} and true value depending on (seeded) extent of error in information requests $\theta_{Z;RL}$ (left panel) and extent of error in action choices (right panel).

6.1 Recovering the data generating process from simulated data

We test how well our estimation procedure can recover the type distribution from simulated data. We use two different type compositions for this purpose. We first ask whether we can recover the underlying data generating process if the true composition is similar to the estimated shares from Section 5.1 with two rules, *RL* and *MBR*. To this aim, we assume that 57% of participants are *RL* and 43% are *MBR* in all simulations (*Exercise 1*) mimicking the population composition estimated in the case of treatment $N - 1$. As a second exercise (*Exercise 2*), we ask how well our procedure performs with three rules in the population. Including three rules also provides a test for whether there is a general tendency for our procedure to favour *RL* and *MBR*, selected in most of our models. To this aim, we simulate the behaviour of three different learning types, *RL* (15% of the population), *MBR* (40%) and *FL* (45% of the population). The specific frequencies for *Exercise 2* were picked in an ad hoc manner with the idea in mind, however, to have a low frequency of *RL* and a higher frequency of *FL* types than what we find in our main estimates.

To mimic our experiment, we simulate data for five groups of eight players (40 computer-simulated “participants” in total) and randomly distribute learning types across computer-simulated “participants”. For each case we have 500 computer-generated samples with these characteristics. We then conduct two different types of simulations. In this Section, we report on simulations where we vary noise in one dimension only, i.e. either in information requests or in action choices, assuming very little noise in the other dimension. Noise parameters θ and ϵ are separately varied between 0 and 1 in steps of 0.2 (keeping the other fixed at 0.05). In all cases, our selection algorithm correctly identifies the rules present in the population.

Figure 8 shows the extent of bias (the difference between estimated and true parameter p_k) as a function of noise in information requests (left panel) and action choices (right panel). The solid line refers to *Exercise 1* (two types) and the dashed line to *Exercise 2* (three types). With noise only in information requests, the bias is near zero for $\theta_Z \lesssim 0.2$, estimates are moderately biased for $0.2 < \theta_Z \lesssim 0.8$ and severely biased for $\theta_Z \gtrsim 0.8$. In our main estimates (Tables 4 and 5) values of θ_Z are generally in the accurate range $\theta_Z \gtrsim 0.2$ with the exception of $N - 1$ where the estimated value is 0.56 suggesting that p_{RL} may be somewhat underestimated in that network. With noise only in action choices, we find that estimates are accurate until values of $\epsilon \approx 0.4$ and imply moderate biases afterwards. If information requests are not very noisy ($\theta_Z = 0.05$), then action requests can be pure noise ($\epsilon = 1$) and the biases will still only be around 10 percentage points. In Appendix B.6 we report on simulations where we simultaneously vary noise in action choices and information requests. Those simulations show that even if we vary noise along both dimensions, biases are small

as long as noise levels are not “too high”. Overall these results make us confident that our estimation procedure works well.

By comparing the left and right panel of Figure 8, we can also say something about how detrimental noise in information requests is as opposed to noise in action choices for recovering the true parameters. This comparison shows that noise in information requests is more likely to lead to biased estimates. In fact, action choices can be complete noise ($\epsilon = 1$), but as long as information requests are accurate, the bias will remain below 15 percentage points. By contrast, if information requests are pure noise ($\theta_Z = 1$), then the fact that there is little noise in action choices, does not help to contain the bias. In this case the frequency of reinforcement learning is severely overestimated, a problem pointed out by Wilcox (2006). These results highlight the importance of incorporating good quality data on information requests in the analysis.

6.2 Alternative Learning Rules

In this section we report results from estimating our model (3) with different variations of belief learning. Under all these variations participants form beliefs based on a fixed number of past periods. 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, since our game lasts only 20 periods. Myopic best responders, by contrast, base their decisions on the last period only. Denote by FP_s the variation, under which participants form beliefs based on the last s periods. Hence, under this terminology, the myopic best-response rule is denoted by 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.

In all treatments the best-performing model is the benchmark from Section 5.1 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 Compliance with Occurrence

In this section we discuss three variations on our assumption of compliance with occurrence.

Easier compliance. Our first variation is more permissive on how often participants should request the minimal information required by each rule in order to classify as compliant. This could be important, if participants consult information less frequently for example after convergence (Section 3). Under the first variation we then assume that a participant has the information she needs to identify the action choice corresponding to her type if she requested the minimal information set *at least once in the preceding four periods*. The estimates are in line with the main results reported in Section 4.2. In $N - 1$ and $N - 2$ we have evidence in favor of RL and MBR ; their estimated population shares are about the same as in the benchmark model in $N - 1$, while in $N - 2$ \hat{p}_{RL} is

higher compared to our benchmark model. In $N - 3$ we now have evidence in favor of *MBR* and *FL*. In all cases the estimated θ 's and ε 's are well behaved (Table 15 in Appendix B.5).

Knowledge about Network Structure. The second variation focuses on observations where a participant has not yet requested information about the network structure and hence cannot (by design) request information about action choices and payoffs of these neighbours. This is the case for less than 10% of our observations. To evaluate to which extent this affects the results we then classify all observations where this is the case as compliant with all rules. The estimated frequencies are unchanged for $N - 1$ and $N - 2$ (compared to Table 4). The same is true for $N - 3$ where all participants end up classified as MBR learners. The rule surviving prior to the last step of elimination is now reinforcement learning rather than forward looking learning as in Table 4 (Table 17 in Appendix B.5).

Number of Compliance Categories. Third, we analyze how sensitive our findings are to the number of compliance categories. Instead of two, we consider three compliance categories under this variation. For each subject i and learning type $k \in \{1, 2, \dots, K\}$, we compute the percentage of times subject i asked for the minimum information required for learning rule k and sort these percentages into three categories: (i) *Z*, 0 compliance with Occurrence; (ii) *M*; 1%-49% compliance with Occurrence; (iii) *H*; 50%-100% compliance with Occurrence. We then estimate a model similar to (3) with $\theta_{kZ} + \theta_{kM} + \theta_{kH} = 1$ for each k .

The results generally confirm our estimates of Section 4.2: *RL* and *MBR* are selected in treatments $N - 1$ and $N - 2$, and only *MBR* in $N - 3$ (Table 16 in Appendix B.5). Why do we prefer the two-level compliance model (3) in Section 4.2 to this one? Notice that in all treatments and for all learning types in Table 16, at least one compliance level (θ_{kj}) is estimated as exactly zero. This indicates that the true number of compliance levels is lower than three and estimating the three-compliance variation may result in identification problems, since (if $\theta_{kM} = 0$) it is possible to find a set of parameters Ω such that all $(p, \theta, \varepsilon) \in \Omega$ generate the same distribution $(\hat{p}, \hat{\theta}, \hat{\varepsilon})$ as in Table 16.

6.4 Alternative Assumptions on Data Generating Process

Poisson distribution. In this subsection we evaluate some alternative assumptions on the data generating process. First, we re-estimate the model assuming that the data-generating process (information requests and choices) follows a Poisson distribution. Let I_k^i denote the number of periods in which subject i searches information consistent with learning type k during the experiment and x_k^i denotes the number of periods 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 may vary across learning rules. Again, in $N - 1$ and $N - 2$ the population is best described by a mix of reinforcement and myopic best response learners, while in $N - 3$ myopic best response and forward looking learning describe the population best (Table 14 in Appendix B.4).

Auto-Correlation. It is well known that it can be important to explicitly account for the panel structure of data when estimating learning models (Frechette, 2009). We hence study an alternative process where we allow for autocorrelation of information requests. We model autocorrelation as a one unit recall Markov process and show that there is some evidence for autocorrelation especially in early periods (1-5) and for own payoff requests, but not for requests in later periods and for

information associated with best response learning. Despite there being some evidence for autocorrelation (especially for RL and early periods), the estimates are roughly similar to those reported in Tables 4 and 5. In fact, for $N - 1$ the estimates are exactly identical and they are nearly identical for $M - 1$ (compared to the case without autocorrelation). For $N - 2$, $N - 3$ and $M - 3$ the share attributed to reinforcement learning is higher with autocorrelation and for $M - 2$ it is lower compared to Tables 4 and 5. In all cases, the two rules selected are reinforcement and myopic best response learning as in the case without autocorrelation (Tables 12 and 13 in Appendix B.3).

6.5 Coordination Game

So far all our results were based on Anti-Coordination games. This raises the question to which extent the results could be game-specific. One might conjecture, for instance, that imitation learning might be more prominent in Coordination Games where neighbours have incentives to choose the same rather than different actions. Our last result in this Section refers to treatments conducted using a Coordination rather than an Anti-Coordination game to get a better sense to which extent the results are game-specific. In three treatments participants played the Coordination game shown in Figure 9 on our networks with intermediate information, i.e. as M -treatments. The bilateral one-shot game has four pure strategy Nash equilibria on the diagonal (i.e. (A,A), (B,B), (C,C) and (D,D)). (A,A) is efficient. Choosing action D has no risk since it guarantees a payoff of 20. As such D is also the maxmin choice. Action B, however, maximizes payoffs against a uniform opponent and is in that sense risk-dominant.

	A	B	C	D
A	60, 60	10, 40	0, 20	0, 20
B	40, 10	40, 40	10, 10	10, 20
C	20, 0	10, 10	30, 30	20, 20
D	20, 20	20, 20	20, 20	20, 20

Coordination Game.

Figure 9: The (bilateral) Coordination Game used in the Experiment.

In all networks, reinforcement learning and myopic best response learning are the two surviving rules. Despite the fact that imitation seems much more intuitive in the Coordination game, there is again no evidence of imitation learning. In contrast to the Anti-Coordination treatments, in the last stage of the algorithm MBR is now very imprecisely estimated with a high θ_Z . One possible reason could be that participants converge faster to equilibrium in the Coordination game and request little information afterwards. Indeed, we do see this in the data. The percentage of successful coordination is 90% in $M-Co-1$, 87% in $M-Co-2$ and 88% in $M-Co-3$ across all periods. Since many participants will stop requesting information after the network has been coordinated for a few periods, identification will be harder after convergence. If we estimate the Coordination Game for the first ten periods only where there is still somewhat more miscoordination we find evidence for only reinforcement learning in $M-Co-1$, evidence for 49% of RL ($\theta_Z = 0.21$) and 51% of MBR ($\theta_Z = 0.78$) in $M-Co-2$ and 45% of RL ($\theta_Z = 0.14$) and 55% of MBR ($\theta_Z = 0.84$) in $M-Co-3$ (Table 25 in Appendix D). Hence even in this case estimates are still quite imprecise. In sum, these results suggest that our findings are not specific to Anti-Coordination Games. They further underline the

importance of choosing a setting where enough off equilibrium choices can be observed, as too quick convergence can lead to imprecise estimates.

7 Concluding Remarks

We use maximum-likelihood methods to estimate learning types in network games. Our estimates are based on both knowledge about which actions participants choose and which information they request in order to make their choice. The latter turns out to be crucial, as estimates that disregard information requests display substantial biases. We also find that network position affects how people learn. In particular those with fewer network neighbours tend to resort to more complex rules. Finally, we find that, while there is substantial heterogeneity in the way people learn in our data, most can be classified as either reinforcement learners or belief learners. Future research is needed to explore the question of heterogeneity and context stability across other dimensions, and within rather than between subjects. Finally, it would be interesting to see how personality characteristics lead people to adopt one or the other learning type as in Gill and Prowse (2015).

References

- [1] Alós-Ferrer, C. and S. Weidenholzer (2008), Contagion and Efficiency, *Journal of Economic Theory* 143, 251-274.
- [2] Apesteguia, J., S. Huck and J. Oechssler (2007), Imitation: Theory and Experimental Evidence, *Journal of Economic Theory* 136: 217-235.
- [3] Arad, A. and A. Rubinstein (2012), Multi-dimensional iterative reasoning in action: the case of the Colonel Blotto game, *Journal of Economic Behavior and Organization* 84, 571-585.
- [4] Blume, L. (1993), The statistical mechanics of strategic interaction, *Games and Economic Behavior*, 5, 387-424.
- [5] Blume, L. (2004), “Evolutionary Equilibrium with Forward-Looking Players”, working paper Santa Fe Institute.
- [6] Börgers, T. and R. Sarin (1997), Learning through Reinforcement and Replicator Dynamics, *Journal of Economic Theory* 77, 1-14.
- [7] Bramoulle, Y. (2007), Anti-Coordination in Networks, *Games and Economic Behavior* 58, 30-49.
- [8] Brock, W.A. and C. Hommes (1997), A Rational Route to Randomness, *Econometrica* 65(5), 1059-1095.
- [9] 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.
- [10] Camerer, C. (2003), *Behavioural Game Theory: Experiments in Strategic Interaction*, Princeton University Press.
- [11] Camerer, C. and Ho, T.-H. (1998), Experienced Weighted Attraction Learning in coordination games: probability rules, heterogeneity, and time-variation, *Journal of mathematical psychology* , 42(2), 305-326.
- [12] Camerer, C. and Ho, T.-H. (1999), Experienced Weighted Attraction Learning in Normal-Form Games, *Econometrica*, 67, 827-874.

- [13] 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.
- [14] Cameron, A. C., and Trivedi, P. K. (2005), *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press.
- [15] Cameron, A. C., and Trivedi, P. K. (2010), "Microeconometrics Using Stata," Stata Press, Ch. 17 p. 599.
- [16] Cheung, Y. and Friedman, D. (1997), Individual Learning in Games: Some Laboratory Results, *Games and Economic Behavior*, 19, 46-76.
- [17] Costa-Gomes, M., Crawford, V., and Broseta, B., (2001), Cognition and Behavior in Normal-Form Games: An Experimental Study, *Econometrica*, 69(5), 1193-1235.
- [18] Efron, B. and Tibshirani, R. J., (1994), *An Introduction to the Bootstrap*, Chapman & Hall/CRC.
- [19] Ellison, G. and D. Fudenberg (1993), Rules of Thumb for Social Learning, *Journal of Political Economy* 101, 612-643.
- [20] 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.
- [21] Eshel, I., L. Samuelson and A. Shaked (1998), Altruists, Egoists and Hooligans in a Local Interaction Model, *American Economic Review* 88, 157-179.
- [22] Feltovich, N. (2000), Reinforcement-Based vs. Belief-Based Learning Models in Experimental Asymmetric-Information Games, *Econometrica* 68(3), 605-641.
- [23] Fischbacher, U. (2007), z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2): 171-178.
- [24] Fosco, C. and F. Mengel (2011), Cooperation through Imitation and Exclusion in Networks, *Journal of Economic Dynamics and Control* 35, 641-658.
- [25] Frechette, G. (2009), Learning in a multilateral bargaining experiment, *Journal of Econometrics* 153, 183-195.
- [26] Fudenberg, D. and D.K. Levine (1998), *The Theory of Learning in Games*, Cambridge: MIT-Press.
- [27] Fujiwara-Greve, T. and C.Krabbe-Nielsen (1999), "Learning to Coordinate by Forward Looking Players", *Rivista Internazionale di Scienze Sociali* CXIII(3), 413-437.
- [28] El-Gamal M.A. and Grether, D.M. (1995), Are People Bayesian? Uncovering behavioural Strategies, *Journal of the American Statistical Association* 90(432), 1137-1145.
- [29] Galeotti, A., Goyal, S., Jackson, M., Vega-Redondo, F., and Yariv, L. (2010), Network games, *Review of Economic Studies*, 77(1), 218-244.
- [30] Gill, D. and V. Prowse (2015), Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy*, in press.
- [31] Goyal, S. (2003), *Learning in networks: a survey*, *Group formation in economics: networks, clubs, and coalitions*, Ed. Demange, G. and Wooders, M. Cambridge University Press.
- [32] Goyal, S. and F. Vega-Redondo (2005), Network formation and social coordination, *Games and Economic Behavior* 50, 178-207.

- [33] Grimm, V. and F. Mengel (2012), An Experiment on Learning in a Multiple Games Environment, *Journal of Economic Theory* 147(6), 2220-2259.
- [34] Ho, T.-H., Wang, X., Camerer, C. (2008), Individual Differences in EWA Learning with Partial Payoff Information, *The Economic Journal* 118, 37-59.
- [35] Hojman, D. and A. Szeidl (2006), Endogenous Networks, Social Games and Evolution, *Games and Economic Behavior* 55 (1), 112-130.
- [36] Hopkins, E. (2002), Two Competing Models of How People Learn in Games, *Econometrica* 70(6), 2141-2166.
- [37] Jackson, M. O. and A. Watts (2002), On the formation of interaction networks in social coordination games, *Games and Economic Behavior* 41, 265-291.
- [38] Jackson, M. O. (2010), *Social and Economic Networks*, Princeton University Press.
- [39] 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.
- [40] Karandikar, R., D. Mookherjee, D. Ray and F. Vega Redondo (1998), Evolving aspirations and cooperation, *Journal of Economic Theory* 80, 292-331.
- [41] Kirchkamp, O. and Nagel, R. (2007), "Naive learning and cooperation in network experiments", *Games and Economic Behavior* 58, 269-292.
- [42] Kirman, A. (1992), "Whom or What does the representative individual represent?", *Journal of Economic Perspectives* 6: 117-136.
- [43] Knoepfle, D.T. and J.T. Wang and C.F. Camerer (2009), Studying Learning in Games Using Eye-Tracking, *Journal of the European Economic Association* 7, 388-398.
- [44] Leroux B. G., (1992), Consistent Estimation of a Mixing Distribution, *The Annals of Statistics* 20, 3, 1350-1360.
- [45] McLachlan, G. J. and D. Peel (2000), *Finite mixture models*, New York: Wiley.
- [46] Mengel, F. (2014), Learning by (limited) forward looking players, *Journal of Economic Behavior and Organization* 108, 59-77.
- [47] Mookherjee, D. and Sopher, B. (1997), Learning and Decision Costs in Experimental Constant Sum Games, *Games and Economic Behavior* 19, 97-32.
- [48] 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.
- [49] 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.
- [50] Salmon, T. (2001), An Evaluation of Econometric Models of Adaptive Learning, *Econometrica* 69(6), 1597-1628.
- [51] Schlag, K. (1998), Why Imitate, and If So, How? A Boundedly Rational Approach to Multi-Armed Bandits, *Journal of Economic Theory* 78, 130-56.
- [52] Skyrms, B. and R. Pemantle (2000), A dynamic model of social network formation, *Proceedings of the National Academy of Sciences* 97, 9340-9346.

- [53] Tang, F-F. (2001), Anticipatory Learning in two-person games: some experimental results, *Journal of Economic Behavior and Organization* 44, 221-232.
- [54] Vega Redondo, F. (1997), Evolution of Walrasian Behavior, *Econometrica*, 375-384.
- [55] Vega-Redondo, F. (2003), *Economics and the Theory of Games*, Cambridge University Press.
- [56] Vega-Redondo, F. (2007), *Complex Social Networks*, *Econometric Society Monograph* 44, Cambridge University Press.
- [57] Wilcox, N., (2005), *Theories of Learning in Games and Heterogeneity Bias: Some Nonsolutions*, Unpublished Manuscript, University of Houston.
- [58] Wilcox, N., (2006), *Theories of Learning in Games and Heterogeneity Bias*, *Econometrica*, 74(5), 1271-1292.

APPENDIX

A Appendix: Detailed Analysis Nash Equilibrium

The table below lists the strict Nash equilibria of the one-shot network game. The entries are vectors (a_1, \dots, a_8) showing the action choices of players 1, ..., 8 as labelled in Section 2.2. The Nash equilibria marked in bold type are also Nash equilibria in every pairwise interaction between first-order neighbours viewed in isolation. This is not true for the other Nash equilibria.

<i>Nash equilibria</i>		
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 8: Strict Nash equilibria. The format is (a_1, \dots, a_8) where $a_i, i = 1, \dots, 8$ is the action of player i .

There are also many non-strict and in particular also mixed equilibria in these games, even though the game with only one neighbour (the bilateral game) has only pure strategy equilibria. We focus predominantly on strict Nash equilibria, since the set of strict Nash equilibria coincides with the set of asymptotically stable outcomes under the multi-population replicator dynamics. The replicator dynamics in turn has been shown to approximate many well known learning models, including reinforcement, fictitious play and variants of imitation learning (see e.g. Hopkins, 2002).

	Network 1	Network 2	Network 3
N-treatments	0.12	0.17	0.00
M-treatments	0.04	0.12	0.00
F-treatments	0.13	0.46	0.00

Table 9: Percentage of successful coordination on a strict Nash equilibrium in last 5 rounds.

Table 9 shows the percentage of successfully coordinated networks in the last 5 periods (on any one-shot Nash equilibrium). Note that coordinating a network of 8 players is difficult and the probability that a network would be coordinated by chance if all players choose actions uniformly at random is ≈ 0.00018 for Network 1 and ≈ 0.00015 for Networks 2 and 3. The table shows that across all information conditions coordination is best in Network 2 and worst in Network 3, where participants never manage to coordinate fully.

Is there more coordination to Nash equilibrium over time? Table 10 shows the results of a logit regression of a binary variable that indicates whether a network was in Nash equilibrium regressed on period 1, ..., 20 as well as a square polynomial of period in odd columns. For each network we pool data across the three information conditions in this table (though the results look similar if we

VARIABLES	(1) Network 1	(2) Network 1	(3) Network 2	(4) Network 2
period	1.393*** (0.463)	0.227*** (0.037)	0.154 (0.113)	0.213*** (0.047)
period \times period	-0.040** (0.017)		0.002 (0.005)	
Constant	-14.09*** (2.86)	-6.18*** (0.97)	-5.05*** (0.55)	-5.38*** (1.01)
Observations	2,080	2,080	2,240	2,240

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10: Convergence to Nash Equilibrium. Logit Regressions of binary variable indicating whether network was in NE on period and square term.

disaggregate). There is no regression for Network 3, because it is never coordinated in any of the treatments. The regression shows that there is more coordination over time in both networks. In Network 1 the square term is negative and statistically significant indicating that learning slows down in later periods, while in Network 2 the square term is close to zero and not statistically significant.

	N-treatments			M-treatments		
	Network 1	Network 2	Network 3	Network 1	Network 2	Network 3
A	0.04	0.05	0.06	0.06	0.07	0.09
B	0.08	0.06	0.07	0.10	0.18	0.13
C	0.36	0.43	0.35	0.32	0.26	0.36
D	0.52	0.46	0.52	0.53	0.49	0.41
	F-treatments					
	Network 1	Network 2	Network 3	Bilateral		
A	0.05	0.00	0.06	0.09		
B	0.13	0.03	0.06	0.25		
C	0.33	0.46	0.37	0.25		
D	0.49	0.51	0.51	0.38		

Table 11: Distribution of Choices in last 10 Periods.

Table 11 shows the distribution of choices in the last 10 periods. Actions C and D are most frequently chosen indicating that players in all networks and treatments attempt to coordinate on equilibria involving these actions.

B Appendix: Additional Background and Results

In this Appendix we collect additional results. Appendix B.1 describes the algorithms behind each learning rule. In Appendix B.2 we compute the correlation coefficient between compliance and action errors. In Appendix B.3 we study a process with auto-correlation. In Appendix B.4 we study a Poisson process. Appendix B.5 collects a variety of robustness checks on the estimates from the N treatments. Finally, in Appendix B.6 we discuss the simulation results in more detail.

B.1 Algorithms Learning Rules

We present the algorithms corresponding to each learning rule. In each period, participants play a 4×4 game against their neighbours and the set of actions is $\{a, b, c, d\}$ for all players.

Under reinforcement learning, participants choose strategies that have performed well in the past with larger probabilities. More precisely, at period t each participant i has a propensity to play each of her four actions. Let $q_i(z, t)$ represent participant i 's propensity at time t of playing action z , for $z \in \{a, b, c, d\}$. These propensities are updated using the updating rule: $q_i(z, t + 1) = q_i(z, t) + \phi$ if i played z 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 period $t + 1$ if

$$q_i(z, t) \in \max\{q_i(a, t), q_i(b, t), q_i(c, t), q_i(d, t)\} \quad (7)$$

In the theoretical literature the choice rule is usually “less deterministic”. Typically players are assumed to choose actions with a probability that is linearly proportional to propensities. The latter assumption means that a larger share of choices can be attributed to reinforcement learning because of the randomization. The second class of learning model we consider is imitation learning. Let N_i^R denote the set of R th order neighbours of i , with $R \in \{1, 2, 3, 4\}$. Under payoff based imitation of order R , learners copy the most successful strategy among their R th order neighbours. Let $\Delta_i^R(z, t)$ represent the average payoff of those players who played action z in period t within participant's i R th order neighbourhood. Player i , then, at time $t + 1$ chooses action z if

$$\Delta_i^R(z, t) \in \max\{\Delta_i^R(a, t), \Delta_i^R(b, t), \Delta_i^R(c, t), \Delta_i^R(d, t)\} \quad (8)$$

Under belief learning models participants 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 participant i gives to her opponents playing each pure strategy z in period t . Therefore player i believes her opponents in period 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 that probability distribution. Under fictitious play agents consider the whole history of the game to compute these probability distribution. Let $Z_i(z, t)$ represent the set of player i 's first order neighbours who played pure strategy z at period t with cardinality $n_i(z, t)$. In the first period fictitious players choose randomly. For all subsequent periods players update their belief vector as follows: $v_i(z, t) = v_i(z, t - 1) + n_i(z, t)$. 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 responders 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 period 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.

B.2 Correlation between Compliance and Action Errors

We ask how errors are correlated. If we define by \mathbf{I}_{kO}^{it} a random variable that equals 1 if participant i searches information consistent with k at period t and define by \mathbf{X}_{kO}^{it} a random variable that equals 1 if participant i chooses an action consistent with k at period t (conditional on compliance O), then we can derive the following correlation coefficient.

From now on we consider a given learning rule k , a given subject i and c possible action choices. Hereafter, we omit subscript k and superscripts i and t . We compute the correlation between $\mathbf{I}_{O(t)}$ and $\mathbf{X}_{O(t)}$.

The conditional probability of taking one of the decisions consistent with k at t is equal to:

$$P(\mathbf{X}_{O(t)} = 1 | \mathbf{I}_{O(t)} = 1) = \left(1 - \frac{4-c}{4}\varepsilon_O\right)\frac{1}{c}, \quad (9)$$

and the probability of choosing a given action that is inconsistent with k is $P(\mathbf{X}_{O(t)} = 0 | \mathbf{I}_{O(t)} = 1) = \frac{\varepsilon_O}{4}$.²⁵ The expected value and the variance of $\mathbf{I}_{O(t)}$ are hence equal to:

$$E[\mathbf{I}_{O(t)}] = 1 - \theta_Z, \quad (10)$$

$$V[\mathbf{I}_{O(t)}] = \theta_Z(1 - \theta_Z). \quad (11)$$

We now compute the conditional expectation of $\mathbf{X}_{O(t)}$:

$$\begin{aligned} E[\mathbf{X}_{O(t)} | \mathbf{I}_{O(t)} = s] &= \sum_{r=0}^1 r \times P(\mathbf{X}_{O(t)} = r | \mathbf{I}_{O(t)} = s), \\ &= P(\mathbf{X}_{O(t)} = 1 | \mathbf{I}_{O(t)} = s), \text{ with } s = \{0, 1\}, \end{aligned} \quad (12)$$

and the expected value of $\mathbf{X}_{O(t)}$ is

$$E[\mathbf{X}_{O(t)}] = \left(\left(1 - \frac{4-c}{4}\varepsilon_O\right)\frac{1}{c}\right)(1 - \theta_Z). \quad (13)$$

To abbreviate notation let $\alpha_O = \left(1 - \frac{4-c}{4}\varepsilon_O\right)\frac{1}{c}$. The unconditional variance of $\mathbf{X}_{O(t)}$ is equal to

$$\begin{aligned} V[\mathbf{X}_{O(t)}] &= E_{\mathbf{I}_{O(t)}}[V[\mathbf{X}_{O(t)} | \mathbf{I}_{O(t)} = s]] + V_{\mathbf{I}_{O(t)}}[E[\mathbf{X}_{O(t)} | \mathbf{I}_{O(t)}]], \\ &= \alpha_O(1 - \theta_Z)\left(\alpha_O\frac{\varepsilon_O}{4} + (1 - \alpha_O)^2 + \alpha_O\theta_Z\right). \end{aligned} \quad (14)$$

Note that if $\theta_Z = 1$, then $V[\mathbf{X}_{O(t)}] = 0$ since $\mathbf{X}_{O(t)}$ is always equal to 0; individuals cannot identify the corresponding action choices simply because they do not have the necessary information to do so.

We now compute the covariance between $\mathbf{I}_{O(t)}$ and $\mathbf{X}_{O(t)}$.

$$\begin{aligned} Cov(\mathbf{I}_{O(t)}, \mathbf{X}_{O(t)}) &= E[\mathbf{I}_{O(t)}\mathbf{X}_{O(t)}] - E[\mathbf{I}_{O(t)}]E[\mathbf{X}_{O(t)}], \\ &= \theta_Z(1 - \theta_Z)\alpha_O > 0, \end{aligned} \quad (15)$$

and the correlation coefficient between information acquisition and action choices is equal to:

$$\rho = \frac{\theta_Z(1 - \theta_Z)\alpha_O}{\sqrt{V[\mathbf{I}_{O(t)}]}\sqrt{V[\mathbf{X}_{O(t)}]}} > 0. \quad (16)$$

This coefficient depends on the parameters θ_Z (probability of making a mistake when searching information consistent with rule k) and ε_O (the conditional probability of making a mistake with compliance O).

²⁵This is similar to CCB, see pp. 1223.

B.3 Autocorrelation Information Requests

We use a Markov chain over two states as a way to model the binary time series \mathbf{I}_{kj}^{it} with $j \in \{Z, O\}$, $k \in \{1, 2, \dots, K\}$ and for $t = \{1, 2, \dots, T\}$. In particular, we assume that the probability that $\mathbf{I}_{kO}^{i;t+1}$ takes value 0 or 1 depends only on the value of the previous outcome \mathbf{I}_{kO}^{it} . The process starts at $t = 1$ with the following probabilities:²⁶

$$\begin{aligned}\Pr(\mathbf{I}_{kZ}^{i1} = 0) &= \Pr(\mathbf{I}_{kO}^{i1} = 1) = 1 - \theta_{kZ}, \\ \Pr(\mathbf{I}_{kZ}^{i1} = 1) &= \Pr(\mathbf{I}_{kO}^{i1} = 0) = \theta_{kZ}.\end{aligned}$$

The transition matrix is equal to:

$$\begin{bmatrix} \pi_{00}^k & 1 - \pi_{00}^k \\ 1 - \pi_{11}^k & \pi_{11}^k \end{bmatrix}$$

where $\pi_{00}^k = P(\mathbf{I}_{kO}^{i;t+1} = 0 | \mathbf{I}_{kO}^{it} = 0)$ and $\pi_{11}^k = P(\mathbf{I}_{kO}^{i;t+1} = 1 | \mathbf{I}_{kO}^{it} = 1)$.

The Markov structure implies that the probability of deviating from information search consistent with k at time $(t + 1)$ equals

$$\theta_{kZ(t+1)} = (1 - \theta_{kZ(t)})(1 - \pi_{11}^k) + \theta_{kZ(t)}\pi_{00}^k$$

After re-arranging terms we have:

$$\theta_{kZ(t+1)} = (1 - \pi_{11}^k) \frac{1 - \Delta^t}{1 - \Delta} + \Delta^t \theta_{kZ}, \quad (17)$$

with $\Delta = (\pi_{00}^k + \pi_{11}^k - 1)$.

The probability of requesting information consistent with k at $(t + 1)$ is then

$$\theta_{kO(t+1)} = \pi_{11}^k \frac{1 - \Delta^t}{1 - \Delta^t} - \frac{\Delta^t(1 - \Delta^{t-1})}{1 - \Delta^t} - \Delta^t \theta_{kZ} \quad (18)$$

The log-likelihood function is

$$\log L(\Psi) = \sum_{i=1}^N \ln \left(\sum_{k=1}^K p_k \left(\prod_{j \in \{Z, O\}} \left(\prod_{t=1}^T \theta_{kjt}^{I_{kjt}^i} \right) \left(\frac{\varepsilon_{jk}}{4} \right)^{(I_{kj}^i - X_{kj}^i)} \prod_{c=1}^4 \left(\left(1 - \frac{4-c}{4} \varepsilon_{jk} \right) \frac{1}{c} \right)^{X_{kj}^{ic}} \right) \right),$$

where now $\theta_{kj(t=1)} = \theta_{kZ}$ for $j = Z$ and $\theta_{kj(t=1)} = 1 - \theta_{kZ}$ for $j = O$; and $\theta_{kZ(t)}$ and $\theta_{kO(t)}$, for $t = \{2, 3, \dots, T\}$ are given by expressions (17) and (18), respectively. Ψ is a vector with the parameters of interest: $(p_k, \theta_{kZ}, \pi_{00}^k, \pi_{11}^k, \varepsilon_{kZ}, \varepsilon_{kO})$ for $k = \{1, 2, \dots, K\}$. We estimate $\log L(\Psi)$ by applying the Expectation Maximization Algorithm (see MacLachlan and Peel, 2000).

If there is no auto-correlation in the binary time series $D_{kj}^{it}(I)$, then we should observe that the probability θ_{kj} remains constant through time. Figure 10 illustrates the evolution of the estimated parameter $\theta_{kZ(t)}$, for $k = \{RL, MBR\}$, during the experiment.

For reinforcement learning, the figure shows that estimated $\theta_{RL;Z}$ are monotonically increasing in $N - 1$ and $N - 2$ (less compliance over time), though by a small amount. In $N - 3$ they are increasing until about period 10, after which there is near zero compliance. For myopic best response learning, the estimated $\theta_{MBR;Z}$ are quickly decreasing during the first 2-3 rounds and constant afterwards.

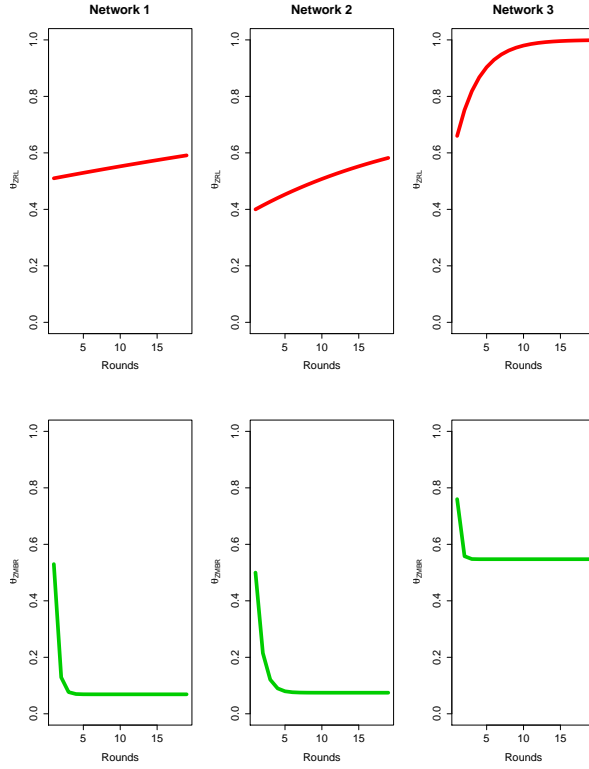


Figure 10: Estimated zero compliance probability θ_{kZ} . Reinforcement learning top panels, myopic best response learning bottom panels. (N -treatments).

	N -Treatments					
	Network 1		Network 2		Network 3	
	RL	MB	RL	MB	RL	MB
$Pr.$ Learning Type	0.57***	0.43***	0.66***	0.34**	0.48***	0.52***
$s.e.$	(0.092)	(0.092)	(0.158)	(0.158)	(0.109)	(0.109)
θ_Z	0.51***	0.53***	0.4*	0.5***	0.66***	0.76***
$s.e.$	(0.179)	(0.167)	(0.205)	(0.174)	(0.127)	(0.117)
π_{00}	1***	0.19	0.99***	0.38	1***	0.57***
$s.e.$	(0.282)	(0.312)	(0.306)	(0.31)	(0.088)	(0.22)
π_{11}	0.99***	0.94***	0.97***	0.95***	0.73**	0.48*
$s.e.$	(0.316)	(0.183)	(0.31)	(0.179)	(0.289)	(0.282)
ϵ_Z	1***	1***	1***	1***	1***	1***
$s.e.$	(0.057)	(0.086)	(0.057)	(0.037)	(0.077)	(0.077)
ϵ_O	0.51***	0.45***	0.51***	0.4***	0.79***	0.58***
$s.e.$	(0.101)	(0.068)	(0.096)	(0.044)	(0.259)	(0.089)
$\bar{\theta}_Z$ (average)	0.55	0.1	0.51	0.1	0.93	0.56
$\bar{\theta}_O$ (average)	0.45	0.9	0.49	0.9	0.07	0.44
$Nb.$ Observations:	760		1064		760	
Loglikelihood:	-1313.48		-1854.79		-1383.45	

Table 12: N -treatments with autocorrelation of order 1 in information requests. Significance levels: (***) 1%, (**) 5% and (*) 10%.

	M-Treatments					
	Network 1		Network 2		Network 3	
	RL	MBR	RL	MBR	RL	MBR
p_k	0.5***	0.5***	0.28***	0.72***	0.22***	0.78***
<i>s.e.</i>	(0.073)	(0.074)	(0.08)	(0.09)	(0.062)	(0.08)
θ_Z	0.05	0.67***	0.1	0.58***	0	0.73***
<i>s.e.</i>	(0.106)	(0.164)	(0.136)	(0.141)	(0.128)	(0.126)
π_{00}	1***	0.99***	0	1***	0.81***	1***
<i>s.e.</i>	(0.316)	(0.084)	(0.33)	(0.085)	(0.300)	(0.080)
π_{11}	0.98***	0.76**	0.72***	0.85***	0.96***	0.88***
<i>s.e.</i>	(0.104)	(0.306)	(0.133)	(0.318)	(0.123)	(0.306)
ϵ_Z	1***	1***	1***	1***	1***	1***
<i>s.e.</i>	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)	(0.077)
ϵ_O	0.58***	0.49***	0.58***	0.51***	0.8***	0.61***
<i>s.e.</i>	(0.072)	(0.136)	(0.121)	(0.116)	(0.109)	(0.118)
$\bar{\theta}_Z$	0.2	0.9	0.21	0.86	0.15	0.89
$\bar{\theta}_O$	0.8	0.1	0.79	0.14	0.85	0.11
Nub. Observations:	760		608		608	
Loglikelihood:	-1295.69		-1060.22		-1047.2	

Table 13: M -treatments with autocorrelation of order 1 in information requests. Significance levels: (***) 1%, (**) 5% and (*) 10%.

Overall, the evidence seems to suggest weak auto-correlation. The patterns are very similar across networks, though the levels differ somewhat.

Tables 12 and 13 show results when allowing for autocorrelation for both the N - and the M -treatments. It can be seen that, despite their being some evidence for auto-correlation (especially for RL and early periods), the estimates are roughly similar to those reported in Tables 4 and 5. In fact, for $N - 1$ the estimates are exactly identical and they are nearly identical for $M - 1$ (compared to the case without autocorrelation). For $N - 2$, $N - 3$ and $M - 3$ the share attributed to reinforcement learning is higher with autocorrelation and for $M - 2$ it is lower compared to Tables 4 and 5.

B.4 Poisson Process

In order to assess 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.

Let I_k^i denote the number of periods in which subject i searches information consistent with learning type k during the experiment and x_k^i denotes the number of periods 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

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

and the log-likelihood function is

$$\ln LF(p, \mu, \lambda | I, x) = \sum_{i=1}^N \ln \left(\sum_{k=1}^K p_k \frac{e^{-\mu_k} \mu_k^{I_k^i} e^{-\lambda_k} \lambda_k^{x_k^i}}{I_k^i! x_k^i!} \right). \quad (19)$$

²⁶Recall that $\mathbf{I}_{kZ}^i = 1 - \mathbf{I}_{kO}^i$.

We apply a selection algorithm similar to that of Section 5. If a learning rule has an estimated μ_k higher than a threshold $\bar{\mu}$ we remove it from the set of rules considered. Table 14 shows the estimation results.

	Endogenous Information Treatments					
	$N - 1$		$N - 2$		$N - 3$	
	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>MBR</i>	<i>FL</i>
p_k	0.58***	0.42***	0.52***	0.48***	0.30***	0.70***
(<i>s.e.</i>)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)
μ_k	1.92	7***	2.15	6.34***	12.76***	0.21
(<i>s.e.</i>)	(3.09)	(2.95)	(2.48)	(2.24)	(2.18)	(0.27)
λ_k	1.02	4.41***	1.15	3.99***	5.68***	0
(<i>s.e.</i>)	(1.83)	(1.95)	(1.54)	(1.58)	(1.85)	(0.14)
ln LL	-255		-339		-114	
Individuals	40		56		40	
Observations	760		1064		760	

Table 14: Poisson distribution. Estimation based on information request and observed behaviour. Note: (***) significant at 1% level, (**) at 5% level and (*) at 10% level. Standard errors computed by bootstrapping methods with 500 replications (Efron and Tibshirani, 1994).

B.5 Robustness Estimates N -treatments

This subsection collects a number of tables with robustness estimates for the N -treatments. Tables 15 and 16 show estimates under alternative assumptions on occurrence. Specifically in Table 15 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 preceding four periods (see Section 6.3). For the estimates reported in Table 16 we allow for three rather than two compliance levels (see Section 6.3).

Parameters	Treatments					
	$N - 1$		$N - 2$		$N - 3$	
	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>MBR</i>	<i>FL</i>
p_k	0.5***	0.5***	0.49***	0.51***	0.43***	0.57***
(<i>s.e.</i>)	(0.087)	(0.087)	(0.076)	(0.076)	(0.091)	(0.090)
θ_Z	0.04*	0.11**	0.11***	0.02	0.21***	0.04
(<i>s.e.</i>)	(0.023)	(0.049)	(0.014)	(0.045)	(0.009)	(0.045)
θ_O	0.96***	0.89***	0.89***	0.98***	0.79***	0.96***
(<i>s.e.</i>)	(0.023)	(0.049)	(0.014)	(0.045)	(0.009)	(0.045)
ε_Z	1***	1***	1***	1***	1***	1***
(<i>s.e.</i>)	(0.045)	(0.045)	(0.045)	(0.045)	(0.051)	(0.045)
ε_O	0.63***	0.44	0.44***	0.51*	0.41***	0.73**
(<i>s.e.</i>)	(0.068)	(0.31)	(0.046)	(0.287)	(0.018)	(0.297)
ln LL	-1071		-1407		-1173	
Individuals	40		56		40	
Observations	760		1064		760	

Table 15: Estimation Results under relaxed assumptions on Occurrence.

We also conduct a robustness check to deal with participants in the N -treatments who do not request information about labels of first-order neighbours (and hence cannot by design request information about their actions or payoffs). Remember that those are relatively few observations. In

Parameters	Treatments					
	$N - 1$		$N - 2$		$N - 3$	
	RL	MBR	RL	MBR	RL	MBR
p_k	0.57***	0.43***	0.59***	0.41***	0.68***	0.32**
(s.e.)	(0.109)	(0.109)	(0.12)	(0.12)	(0.139)	(0.139)
θ_{kZ}	0.55*	0.1	0.48	0.14	0.9***	0.65***
(s.e.)	(0.307)	(0.260)	(0.292)	(0.253)	(0.142)	(0.225)
θ_{kM}	0.06	0	0.09*	0	0.1*	0.11*
(s.e.)	(0.039)	(0.032)	(0.048)	(0.059)	(0.052)	(0.061)
θ_{kH}	0.39	0.9***	0.43	0.85*	0	0.23
(s.e.)	(0.373)	(0.316)	(0.329)	(0.306)	(0.1967)	(0.313)
ε_{kZ}	1***	1***	1***	1***	1***	1***
(s.e.)	(0.046)	(0.045)	(0.045)	(0.045)	(0.045)	(0.064)
ε_{kM}	0.52**	0.9***	0.5***	0.87***	0.58***	0.74***
(s.e.)	(0.202)	(0.263)	(0.171)	(0.248)	(0.178)	(0.229)
ε_{kH}	0.51***	0.46***	0.5***	0.41***	0.47*	0.57***
(s.e.)	(0.123)	(0.136)	(0.118)	(0.134)	(0.240)	(0.15007)
ln LL		-1406		-2019		-1474
Individuals		40		56		40
Observations		760		1064		760

Table 16: Estimation results with three compliance levels. Note: (***) significant at 1% level, (**) significance at 5%, (*) significance at 10% level. Standard errors computed by bootstrapping method with 500 replications (Efron and Tibshirani, 1994).

period 1 already more than 70% request information about first-order neighbours and $\approx 90\%$ do so at some point during the experiment.

Table 17 shows estimates where these observations are classified automatically as compliant for all rules. The estimated frequencies are unchanged for $N - 1$ and $N - 2$ (compared to Table 4). The same is true for $N - 3$ where all participants end up classified as MBR learners. The rule surviving prior to the last step of elimination is now reinforcement learning rather than forward looking learning as in Table 4.

B.6 Simulations

This subsection contains additional information, tables and results regarding the simulations reported on in Section 6.1. In addition to the simulations described in the main text we also ran simulations, where we simultaneously increase noise in information requests and noise in action choices. We use the same type distributions as in Exercise 1 in Section 6.1. For each parameter combination we ran 500 replications. We first consider three different parameter constellations as follows:

1. *Full Compliance (FC)*: participants search their respective information set with probability 1 and make no mistake in choosing the corresponding action choice.
2. *High Compliance (HC)* hereafter): participants search their corresponding information request with high probability and make mistakes with low probability,
3. *Low Compliance (LC)*: participants have low compliance with occurrence and make mistakes with high probability.

Table 18 reports the results. Our selection algorithm always correctly identifies the learning rules present in the population. The shares of *PBI* and *FL* are virtually zero in the two-type case; the same holds for *PBI* in *Exercise 2*. Moreover, we find only small biases (less than 1%) in the estimated frequencies in both **FC** and **HC**. Hence, if people are relatively precise both making their choices and looking up the information, our estimation procedure succeeds in recovering the population composition in all cases. As participants become less precise in their information requests

Parameters	Treatments					
	M-1		M-2		M-3	
	RL	MBR	RL	MBR	RL	MBR
p_k	0.57***	0.43***	0.24	0.76***	0.47***	0.53***
(s.e.)	(0.100)	(0.100)	(0.167)	(0.167)	(0.118)	(0.118)
θ_Z	0.56*	0.10	0.05	0.47**	0.93***	0.60**
(s.e.)	(0.307)	(0.267)	(0.283)	(0.208)	(0.200)	(0.298)
θ_O	0.44	0.90**	0.95***	0.53**	0.07	0.40
(s.e.)	(0.307)	(0.267)	(0.283)	(0.208)	(0.200)	(0.298)
ε_Z	1***	1***	1***	1***	1***	1***
(s.e.)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
ε_O	0.51***	0.45***	0.48***	0.43***	0.49***	0.59***
(s.e.)	(0.088)	(0.113)	(0.07)	(0.068)	(0.186)	(0.148)
ln LL	-1325		-1861		-1398	
Individuals	40		56		40	
Observations	800		1120		800	

Table 17: Main Estimation Results using *both* information requests and action choices. N-treatments. All observations where a participant does not know anything about the network structure are automatically classified as compliant. Note: (***) significant at 1%, (**) at 5% and (*) at the 10% level. Standard errors computed using bootstrapping method with 500 replications following Efron and Tibshirani (1994).

and decisions (**LC**), we still recover which types are present in the population, but there are biases in the estimated values. In the two-type case, the mechanism overestimates the presence of *MBR* by 11% and underestimates the share of *RL* by 16% (4% and 1% of the biases correspond to *PBI* and *FL*, respectively; though they are not significantly different from zero). Similar conclusions hold for *Exercise 2*.

We then increase noise simultaneously in steps of 0.1. Figure 11 shows the distribution of estimated parameters \hat{p}_{RL} across the 500 runs for *Exercise 1*. The leftmost panel shows the case where $(\theta_Z, \varepsilon_O) = (0.15, 0.1)$. Estimated parameters are concentrated around the true value of 0.575, the standard deviation is 0.01 and the mean bias 0. Estimates at these noise levels are hence extremely accurate. As we increase noise estimates tend towards uniform distribution, which means that, since the “true” value of p_{RL} is above 0.5, the estimates will be downward biased. The middle panel shows the case $(\theta_Z, \varepsilon_O) = (0.55, 0.55)$, here the mean bias is -0.17 , i.e. most simulations converge to a case where the parameter is underestimated. Around 5 percent of simulations only converge to the true value. Further increasing the amount of noise does not substantially worsen the bias as the rightmost panel illustrates, which shows the case $(\theta_Z, \varepsilon_O) = (0.75, 0.70)$.

	Exercise 1				Exercise 2			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
True p_k	0.57	0.00	0.43	0.00	0.15	0.00	0.40	0.45
FC								
\hat{p}_k	0.562	0.014	0.423	0.001	0.150	0.000	0.400	0.450
(s.e.)	(0.09)	(0.04)	(0.09)	(0.004)	(0.001)	(0.000)	(0.002)	(0.001)
Bias	-0.0077	0.014	-0.007	0.0008	-0.001	0.000	0.000	0.000
HC								
\hat{p}_k	0.57	0	0.425	0	0.1500	0.0000	0.4000	0.4500
(s.e.)	(0.0049)	(0)	(0)	(0)	(0.0000)	(0.0000)	(0.0005)	(0.0004)
Bias	0.005	0	-0.005	0	0.0001	0.0000	0.0000	0.0000
LC								
\hat{p}_k	0.404	0.042	0.542	0.012	0.271	0.05	0.323	0.355
(s.e.)	(0.203)	(0.205)	(0.183)	(0.109)	(0.376)	(0.224)	(0.248)	(0.213)
Bias	-0.165	0.041	0.111	0.0119	0.121	0.050	-0.077	-0.095
<i>Assumptions</i>								
θ_{kZ}	0	0.15	0.55		0	0.15	0.55	
θ_{kO}	1	0.85	0.45		1	0.85	0.45	
ε_{kZ}	1	1	1		1	1	1	
ε_{kO}	0	0.1	0.55		0	0.1	0.55	

Table 18: Monte Carlo Simulations.

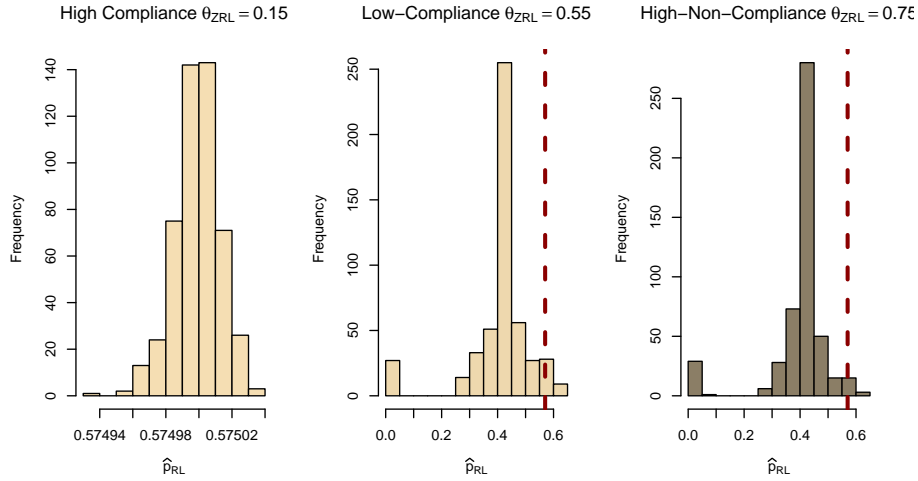


Figure 11: Distribution of estimated parameters \hat{p}_{RL} across 500 runs of simulations of Exercise 1 (note the different scale on x-axis across panels).

C Appendix: Additional Figures

Figure 12 provides an overview of information requests over time in the N -treatments. Figure 13 shows the analogous information for the M -treatment. Figure 14 shows the fractions of participants who *both* request the minimal necessary information corresponding to a rule and choose as prescribed by that rule.

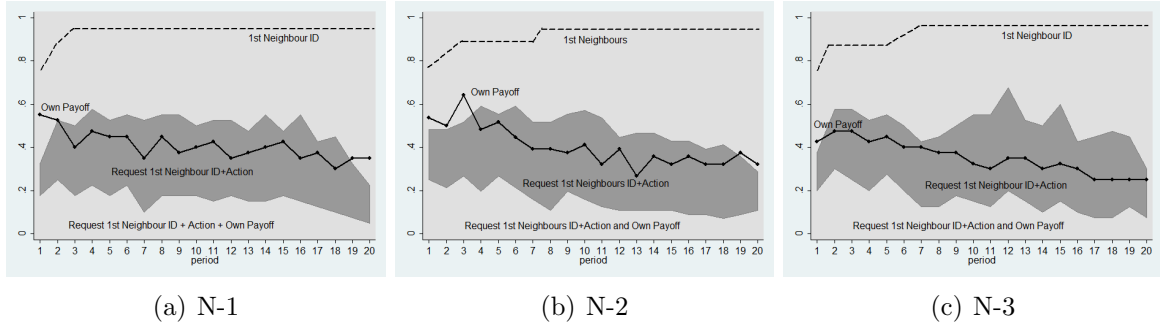


Figure 12: Share of participants (i) requesting information about all three: 1st neighbours ID (experimental label)+actions as well as own payoffs (bottom area), (ii) requesting only 1st order neighbours ID and actions, but not payoffs (dark area), (iii) requesting own payoffs (black solid line) and (iv) requesting only own payoff (area between lower bound of dark area and black line). N-treatments

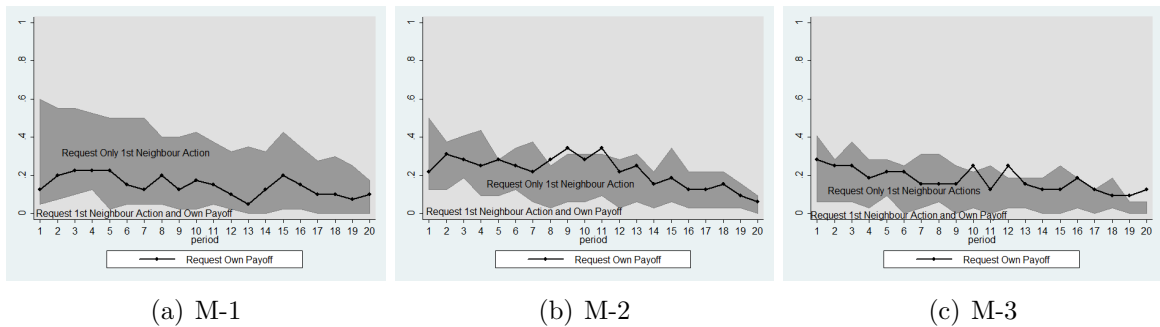


Figure 13: Share of participants requesting information about all three: 1st neighbours actions as well as own payoffs (bottom area), requesting only 1st order neighbours actions, but not payoffs (dark area), requesting own payoffs (black curve) and requesting only own payoff (area between lower bound of dark area and black curve). M-treatments

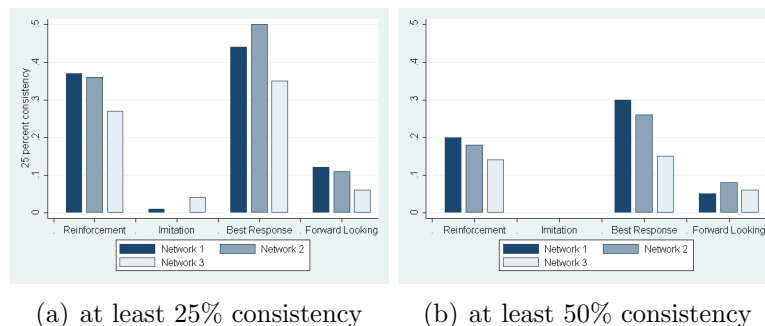


Figure 14: Fraction of participants who request the minimal information set *and* play the action prescribed by the corresponding learning type more than 25% (left) or 50% (right) of periods.

D Appendix: Additional Tables

Table 19 shows how often different rules predict different choices conditional on the histories in our experiment and Table 20 shows the minimal information required for each rule.

Learning Rules	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

Learning Rules	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

Learning Rules	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 19: Separation between learning types on basis of action choices. Each cell contains the average number of periods in which the two corresponding learning types predict different choices for a participant (given the history of play in the experiment).

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

Table 20: Minimal Information Required for Each Rule (x indicates that a piece of information is required for the corresponding learning rule).

Table 21 shows the results separately for periods 1-10 and 11-20 for the N -treatments. Results seem overall robust. The two surviving rules are always reinforcement learning and/or myopic best responses. In $N - 1$ reinforcement learning has a share of 59% across periods 1-10 and of 63% across periods 11-20 with the remainder being attributed to MBR. In $N - 2$ reinforcement learning receives a share of 55% across periods 1-10 and 73% across periods 11-20 and in $N - 3$ MBR is the unique surviving rule in both cases. Hence in $N - 1$ and $N - 3$ estimates are very similar across the two blocks and also very similar to the estimates based on all 20 periods. In $N - 2$ the estimates based on subperiods differ somewhat from the estimates using all periods attributing a higher share to RL. One possible reason for this difference seems to be that lots of noisy decisions are attributed to RL in this network, as indicated by relatively high values of $\theta_{RL;Z}$ in the estimations based on subperiods.

Table 22 shows results for splitting the N -treatment data by whether participants have one or more than one network neighbour. The difference to the results reported in Table 7 in the main text is that for the estimates here network $N - 1$ data are included (in Group 2). Table 23 reports the same results for the M -treatments. Table 24 shows estimates using only choice behaviour in the F -treatments. Table 25 reports the estimates from our Coordination Game treatments.

	<i>Treatment N1</i>				<i>Treatment N2</i>				<i>Treatment N3</i>			
	<i>Periods 1 to 10</i>		<i>Periods 11 to 20</i>		<i>Periods 1 to 10</i>		<i>Periods 11 to 20</i>		<i>Periods 1 to 10</i>		<i>Periods 11 to 20</i>	
	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>	<i>RL</i>	<i>MBR</i>
p_k	0.59***	0.41***	0.63***	0.37***	0.55***	0.45***	0.73***	0.27**	0.31***	0.69***	0.65***	0.35***
(s.e.)	(0.109)	(0.109)	(0.116)	(0.115)	(0.117)	(0.117)	(0.193)	(0.193)	(0.148)	(0.148)	(0.101)	(0.101)
θ_Z	0.48*	0.07	0.57*	0.04	0.41	0.11	0.62*	0.02	0.99***	0.55***	0.96***	0.62**
(s.e.)	(0.256)	(0.338)	(0.291)	(0.268)	(0.190)	(0.159)	(0.303)	(0.281)	(0.216)	(0.223)	(0.232)	(0.263)
θ_O	0.52**	0.93*	0.43*	0.96***	0.59***	0.89***	0.38	0.98***	0.01	0.45*	0.04	0.38
(s.e.)	(0.256)	(0.338)	(0.291)	(0.268)	(0.190)	(0.159)	(0.303)	(0.281)	(0.216)	(0.223)	(0.232)	(0.263)
ε_Z	1***	1***	1***	1***	1***	1***	1***	1***	1***	1***	1***	1***
(s.e.)	(0.045)	(0.044)	(0.071)	(0.059)	(0.045)	(0.045)	(0.063)	(0.057)	(0.045)	(0.045)	(0.045)	(=,0.045)
ε_O	0.69***	0.58***	0.3*	0.35***	0.61***	0.49***	0.32*	0.32***	1	0.68***	0	0.51***
(s.e.)	(0.125)	(0.154)	(0.141)	(0.119)	(0.062)	(0.069)	(0.138)	(0.093)	(0.296)	(0.174)	(0.157)	(0.218)
$\ln LL$	-663.36		-581.75		-903.09		-833.1		-686.96		-617.02	
<i>Individuals</i>	40		40		56		56		40		40	
<i>Observations</i>	360		360		504		504		360		360	

Table 21: Estimations separately for periods 1-10 and 11-20. N - treatments.

Parameters	Group 1: one network neighbor				Group 2: multiple network neighbors			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.54***		0.46***		0.61***		0.39***	
(s.e.)	(0.10)		(0.10)		(0.147)		(0.147)	
θ_Z	0.87***		0.74*		0.58*		0.11	
(s.e.)	(0.26)		(0.33)		(0.327)		(0.203)	
θ_O	0.13		0.26		0.42		0.89***	
(s.e.)	(0.26)		(0.33)		(0.327)		(0.203)	
ε_Z	1***		1***		1***		1***	
(s.e.)	(0.04)		(0.04)		(0.045)		(0.045)	
ε_O	0.63***		0.7***		0.5***		0.41***	
(s.e.)	(0.12)		(0.11)		(0.063)		(0.063)	
$\ln LL$			-163				-2981	
<i>Individuals</i>			46				90	
<i>Observations</i>			874				1710	

Table 22: 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 (1-8 in N-1, 2,3,4,5,6 in N-2 and 2,3,4 in N-3). Note: (***) significant at 1% level, (**) at 5% level and (*) at 10% level. Standard errors computed by bootstrapping methods with 500 replications (Efron and Tibshirani, 1994).

Parameters	Group 1: one network neighbor				Group 2: multiple network neighbors			
	RL	PBI	MBR	FL	RL	PBI	MBR	FL
p_k	0.23***		0.77***		0.44***		0.56***	
(s.e.)	(0.09)		(0.09)		(0.068)		(0.07)	
θ_Z	0.21		0.86***		0.21***		0.91***	
(s.e.)	(0.149)		(0.053)		(0.04)		(0.051)	
θ_O	0.79***		0.14***		0.79***		0.09*	
(s.e.)	(0.149)		(0.053)		(0.04)		(0.051)	
ε_Z	1***		1***		1***		1***	
(s.e.)	(0.045)		(0.045)		(0.045)		(0.045)	
ε_O	0.68***		0.64***		0.61***		0.44***	
(s.e.)	(0.155)		(0.115)		(0.058)		(0.101)	
$\ln LL$			-1091				-2374	
<i>Individuals</i>			32				72	
<i>Observations</i>			640				1440	

Table 23: Participants with one network neighbour (1,7,8 in M-2 and 1,5,6,7,8 in M-3) vs participants with multiple network neighbours (all players M-1, players 2,3,4,5,6 in M-2 and 2,3,4 in M-3). Note: (***) significant at 1% level, (**) at 5% level and (*) at 10% level. Standard errors computed by bootstrapping methods with 500 replications (Efron and Tibshirani, 1994).

<i>Treatment F - 1</i>				
	<i>RL</i>	<i>PBI</i>	<i>MBR</i>	<i>FL</i>
p_k	0.52**	0.10	0.34***	0.04
(s.e.)	(0.168)	(0.094)	(0.123)	(0.123)
ε_k	0.53**	1	0.37***	0
(s.e.)	(0.224)	(0.193)	(0.095)	(0.255)
<i>ln LL</i>			-489	
<i>Individuals</i>			24	
<i>Observations</i>			456	
<i>Treatment F - 2</i>				
	<i>RL</i>	<i>PBI</i>	<i>MBR</i>	<i>FL</i>
p_k	0.12	0.21**	0.58***	0.08
(s.e.)	(0.093)	(0.093)	(0.117)	(0.065)
ε_k	0.58	0.68***	0.26***	0
(s.e.)	(0.304)	(0.095)	(0.062)	(0.286)
<i>ln LL</i>			-400	
<i>Individuals</i>			24	
<i>Observations</i>			456	
<i>Treatment F - 3</i>				
	<i>RL</i>	<i>PBI</i>	<i>MBR</i>	<i>FL</i>
p_k	0.32**	0.04	0.34**	0.30**
(s.e.)	(0.124)	(0.062)	(0.144)	(0.118)
ε_k	0.34**	0	0.48***	0.57***
(s.e.)	(0.163)	(0.367)	(0.108)	(0.057)
<i>ln LL</i>			-467	
<i>Individuals</i>			24	
<i>Observations</i>			456	

Table 24: Estimations based solely on observed behavior in F -treatments. Note: (***) Significant at 1% level; (**) at 5% level; (*) at 10% level. Standard errors computed by bootstrapping method with 500 replications (Efron and Tibshirani, 1994)

Parameters	Treatments					
	M-Co-1		M-Co-2		M-Co-3	
	RL	MBR	RL	MBR	RL	MBR
p_k	0.54***	0.46***	0.48***	0.52***	0.68***	0.32**
(s.e.)	(0.105)	(0.105)	(0.095)	(0.095)	(0.139)	(0.139)
θ_Z	0.31***	0.95***	0.27***	0.9***	0.34***	0.92***
(s.e.)	(0.069)	(0.052)	(0.072)	(0.049)	(0.109)	(0.081)
θ_O	0.69***	0.05	0.73***	0.1**	0.66***	0.08
(s.e.)	(0.069)	(0.052)	(0.072)	(0.049)	(0.109)	(0.081)
ε_Z	1***	1***	1***	1***	1***	1***
(s.e.)	(0.06)	(0.00)	(0.04)	(0.04)	(0.04)	(0.04)
ε_O	0.37***	1***	0.38***	1***	0.43***	1***
(s.e.)	(0.072)	(0.116)	(0.098)	(0.05)	(0.057)	(0.059)
<i>ln LL</i>		-1015		-1039		-1335
<i>Individuals</i>		32		32		40
<i>Observations</i>		640		640		800

Table 25: Main Estimation Results using *both* information requests and action choices. M-treatments and Coordination Game. Note: (***) significant at 1%, (**) at 5% and (*) at the 10% level. Standard errors computed using bootstrapping with 500 replications following Efron and Tibshirani (1994).