Breitmoser Y, Tan JHW, Zizzo DJ.

[On the beliefs off the path: Equilibrium refinement due to quantal response and level-K.](#)

Games and Economic Behavior 2014, 86, 102-125.

Copyright:

©2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC-BY license (<http://creativecommons.org/licenses/by/3.0/>)

DOI link to article:

<http://dx.doi.org/10.1016/j.geb.2014.03.002>

Date deposited:

14/04/2016

This work is licensed under a [Creative Commons Attribution 3.0 Unported License](#)

On the beliefs off the path: Equilibrium refinement due to quantal response and level- k

Yves Breitmoser^a, Jonathan H.W. Tan^{b,*}, Daniel John Zizzo^c

^a School of Business and Economics, Humboldt-Universität Berlin, Spandauer Straße 1, 10099 Berlin, Germany

^b Nottingham University Business School, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, UK

^c School of Economics and CBESS, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

ARTICLE INFO

Article history:

Received 29 September 2010

Available online 25 March 2014

JEL classification:

C72

C91

D62

Keywords:

Incomplete information

Equilibrium refinement

Logit equilibrium

Rationalizability

Quantal response

Level- k

Inequity aversion

Experiment

ABSTRACT

The extensive form game we study has multiple perfect equilibria, but it has a unique limiting logit equilibrium (QRE) and a unique level- k prediction as k approaches infinity. The convergence paths of QRE and level- k are different, but they converge to the same limit point. We analyze whether subjects adapt beliefs when gaining experience, and if so whether they take the QRE or the level- k learning path. We estimate transitions between level- k and QRE belief rules using Markov-switching rule learning models. The analysis reveals that subjects take the level- k learning path and that they advance *gradually*, switching from level 1 to 2, from level 2 to equilibrium, and reverting to level 1 after observing opponents deviating from equilibrium. The steady state therefore contains a mixture of behavioral rules: levels 0, 1, 2, and equilibrium with weights of 2.9%, 16.6%, 37.9%, and 42.6%, respectively.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

Current theories of bounded rationality explain behavior in strategic interactions by specifically modeling the noisiness of payoff maximization and the formation of non-equilibrium beliefs. In quantal response equilibrium (QRE, McKelvey and Palfrey, 1995, 1998), players correctly anticipate their co-players' noisy strategies and they respond noisily. Non-equilibrium models, such as level- k reasoning (Stahl and Wilson, 1994; Nagel, 1995; Ho et al., 1998; Costa-Gomes et al., 2001) or cognitive hierarchy (Camerer et al., 2004; Rogers et al., 2009), additionally model how players form conjectures about their co-players' strategies and thereby deviate from equilibrium. Non-equilibrium models have been shown to explain "initial play" in a variety of environments. Examples include auctions (Crawford and Iriberri, 2007) and information cascades (Kübler and Weizsäcker, 2004). In turn, equilibrium models are considered to fit choices of experienced players. Existing studies apply static models. Thus, whether and how subjects actually switch between non-equilibrium and equilibrium beliefs is not yet understood. The present paper seeks to fill this void. We analyze transitions between these two fundamental belief regimes with dynamic models that consider transitions in the cognitive hierarchy.

* Corresponding author. Fax: +44 (0)115 846 6667.

E-mail addresses: yves.breitmoser@hu-berlin.de (Y. Breitmoser), jonathan.tan@nottingham.ac.uk (J.H.W. Tan), d.zizzo@uea.ac.uk (D.J. Zizzo).

To this end, we conduct a laboratory experiment on an extensive form “club game” where the limit points of both QRE and level- k , i.e. at $\lambda, k \rightarrow \infty$, uniquely restrict the equilibrium actions. The main alternative refinement concepts, such as perfection (Selten, 1975), properness (Myerson, 1978), and sequentiality (Kreps and Wilson, 1982), induce a multiplicity of equilibria.¹ The convergence paths differ between level- k and QRE, but they converge to the same limit point—i.e. the mode of reasoning affects the convergence path, but not equilibrium selection. Thus, the club game allows us to distinguish level- k and QRE paths with neither interference from other equilibrium refinement concepts, nor interference due to equilibrium selection. We then measure the transitions between level- k and QRE using a Markov-switching model.

The club game is an extension of Dixit (2003).² Each player moves once, deciding whether to join the club. Players have heterogeneous preferences for joining a club, and payoffs are structured such that one's incentive to join the club increases with the number of opponents who join the club. Following Dixit, we consider the case where player 1 is better off joining regardless of whether or not the opponents join, player 2 is better off joining if at least one opponent joins, player 3 is better off joining if at least two opponents join, and player 4 is better off joining if three other opponents join. Under common knowledge of rationality, all players will join the club: player 1 surely joins, so player 2 is better off joining; players 1 and 2 surely join, so player 3 is better off joining; players 1, 2, and 3 surely join, so player 4 is better off joining. Conversely, player 4 is better off not joining if only two opponents join, player 3 is better off not joining if only one opponent joins, and player 2 is better off not joining if no opponent joins. Thus, a majority comprising players 2, 3, and 4 may be better off when no club forms at all, i.e. if player 1 would not join. We adapt Dixit's game by introducing incomplete information on the order of moves. A player knows his position in the order of moves and the actions of players with preceding moves, but does not know which co-player has moved or is to move. As we show below, this adaptation separates level- k and QRE learning paths, which we exploit in our analysis.

The club game is chosen, as it is a sharp, yet simple experimental tool for identifying level- k and QRE types. In the club game, only the limits of level- k and QRE serve as suitable refinements, and despite being a binary yes–no game, it is iteratively dominance solvable. Iterative—as opposed to non-iterative—dominance solvability separates the different levels of reasoning, and due to the neutrality with respect to perfection and properness, this allows us to analyze rule learning absent of alternative refinement effects. More generally, the club game is a simple but intuitive representation of strategic decision making in environments where beliefs about the rationality of others are of primary relevance. Of course the club game is not the only game with unique yet different limiting level- k and QRE predictions, without interference of perfection and properness, but it seems hard to construct many intuitive, binary games with this property.

Our four experimental treatments involve a 2×2 design. On the one hand, we vary incentives to individuals to join. This facilitates the identification of belief rules, as it varies the predicted QRE strategies without affecting level- k strategies. On the other hand, we manipulate externalities for other players, which allows us to control for social preferences (social preferences are of potential relevance in all multi-player games). Both our methodology, to use strictly sequential games and Markov-switching models in analyses of rule learning, and our results appear directly applicable to all games with similar relevance of beliefs about rationality. This includes games that are iteratively dominance solvable or constant-sum; it may however plausibly exclude, or anyway be less relevant, for games such as public goods games, trust games or joy-of-destruction games which are non-constant-sum and where social preferences and beliefs about others' social preferences are likely to play more significant roles. Of course, the precise range of applicability of our results is an empirical question that may be addressed in future research.

Our study extends existing work on “rule learning”, i.e. transitions between belief rules, in two ways. First, further to analyses such as the seminal work of Stahl (1996, 2000) and Haruvy and Stahl (2012) on normal form games, we consider learning in an extensive form game that varies parametrically during the course of the experiment. Specifically, the club game is an extensive form game with strictly sequential moves: Players observe previous choices, form conjectures about subsequent movers, and then choose accordingly. In relation to normal form games, this implies that players do not need to form beliefs about their opponents' beliefs about oneself—such second-order beliefs are not required, since others' previous choices are known prior to moving. Thus, the extensive form club game allows us to focus on first-order beliefs more sharply than normal form games.

Other extensive form games that are sufficiently simple to be implemented in laboratories tend to have the disadvantage that equilibrium predictions are either unique and equivalent for all refinement concepts, e.g. in ultimatum games, or not unique for weaker concepts such as Perfect Bayesian equilibria and equilibrium refinement then becomes fairly subtle, e.g. in signaling games. As mentioned, the club game avoids these pitfalls by inducing precise, yet, different predictions for level- k and QRE without interference from alternative refinement concepts. The club game is also neutral to teaching as observed by Hyndman et al. (2012). In their experiment, sophisticated subjects forsake payoff maximization in response to their beliefs in order to influence their opponents' beliefs and actions. In our experiment, the players that are most

¹ Perfect, proper, and sequential equilibria do not restrict actions in various information sets off the equilibrium path in the club game (discussed below). Similar to limiting logit equilibrium, these equilibria are limit points as a tremble variance approaches zero. The difference is that perfect (proper, sequential) trembles do not restrict relative tremble probabilities between players, while limiting logit increases precision in payoff maximization, which imposes an intuitive restriction between players.

² The club game was originally inspired by applications of peer effects due to social ties. “Joining a club” can be interpreted as a metaphor for joining in an activity, e.g. riots (Schelling, 1960) or the usage of alcohol and drugs (Gaviria and Raphael, 2001), or even “ethnification”, i.e. adopting a marker for identifying membership to a certain ethnic group (e.g. wearing a certain color of hat, Kuran, 1998).

enthusiastic about the club could assume the positions of teachers, but they find it optimal to join also in response to their beliefs. Hence, long-run teaching does not contradict short-run payoff maximization in the club game, which allows us to study belief formation absent of teaching effects.

Second, we extend existing work by proposing a Markov-switching rule learning model to understand individual transitions between belief rules.³ Specifically, we lean on the approach taken recently by Haruvy and Stahl (2012), but adapt their assumption that players switch between rules based on strategic experimentation and reinforcement learning. In their model, players experiment with different rules in the short run, and due to reinforcement, they converge to the most successful rule in the long run. Since reinforcement learning applies well to action-based learning within games (Erev and Roth, 1998), this approach constitutes a natural starting point to model rule learning.⁴ Similarly to multinomial logit models, for example, reinforcement models treat actions as *nominal*. In contrast, rules about belief formation are *ordered*: level 0, level 1, 2, ..., up to equilibrium. This motivates a possible difference between action-based learning models and rule-based learning models. For, instead of experimenting with different rules, players may systematically advance through the set of rules. Reinforcement rule learning cannot capture systematic transitions patterns.

Systematic transitions are captured by Markov-switching models, which in turn are straightforward generalizations of finite-mixture models (see e.g. McLachlan and Peel, 2000; Frühwirth-Schnatter, 2006). The latter have become the standard approach to modeling mixed populations containing level- k and equilibrium types (following Stahl and Wilson, 1994, 1995) with each subject having a constant level of reasoning throughout the experiment. In the simplest such generalization, Markov-switching models extend finite-mixture models by adding a Markov transition matrix. Thus, the individual level of reasoning may change over the course of the experiment, and the probabilities of the rules employed in the next period may depend on which rule is employed in this period. For parsimony, and to cleanly test the hypothesis that rule switching is gradual rather than non-gradual, we focus on such simple models. An interesting extension would be to allow for payoff-dependent rule switching described by a Markov model with both stochastic and reinforcement components, which may allow future research to identify the events that trigger rule switching.

In the experiment, we observe the convergence predicted by the limit points of QRE and level- k , which in turn rules out perfection, properness, and sequentiality as primary forces of equilibrium refinement in club games. By looking at the actions prior to convergence, we can distinguish learning along level- k path from levels 0, 1, 2, to equilibrium, and QRE paths from level 0 via increasingly precise QREs to equilibrium. Decisions are “close” to equilibrium, particularly in later stages of the experiment. The subjects who deviate from “joining” tend to be those with the weaker incentives to join, which relates to the results of Battalio et al. (2001) and Goeree et al. (2002), who report an “optimization premium” to a similar effect. Rule learning is indeed best explained by the Markov-switching rule learning model. According to the estimated learning paths, individuals transit systematically across rules over time, from level 1 to level 2, and from level 2 to a high-precision equilibrium. Around 15% of subjects at level 2 and equilibrium switch back to level 1 after each period—i.e. subjects may revise their beliefs “downwards” after observing unexpected non-equilibrium decisions. Subjects take the level- k path rather than the QRE path when adapting their beliefs, i.e. they switch between different levels of reasoning rather than different QRE precisions. The sample tends toward a steady state with a mixture of mostly level 2 and equilibrium types.

To summarize, we show that individual beliefs on the level of reasoning of other players are not constant over time, that Markov-switching rule learning models are an effective way of estimating the systematic transitions along the level- k path, and that the steady state is a mixture of types including boundedly rational ones. This plausibly follows from equilibrium subjects abandoning their equilibrium beliefs and reverting to level 1 after observing play by boundedly rational subjects such as level 1 players, thus perpetuating the existence of level 1 players.

Section 2 defines the club game. Section 3 presents the equilibrium analysis. Section 4 describes and motivates the experimental games, procedure and logistics. Section 5 provides an overview of the main behavioral patterns and how they relate qualitatively to QRE and level- k models of bounded rationality. Section 6 structurally analyzes the rules underlying belief formation and rule learning. Section 7 discusses and concludes.

2. The club game

The set of players is $N = \{1, \dots, n\}$, with typical elements $i, j \in N$. The players move sequentially, in random order, and all possible move sequences have positive probability. They have to decide whether to join the club ($a_i = 1$) or not ($a_i = 0$), knowing only how many players have moved before and how many of them have joined. Information sets are denoted as (i, t, k) , where $i \in N$ is the player to move, $t \in \{1, \dots, n\}$ is his position in the move sequence, and $k \in \{0, 1, \dots, t-1\}$ is the

³ While our assumption that belief formation can be described using general rules, such as level- k or equilibrium, is widely used, see Haruvy and Stahl (2012) and the work cited therein, an alternative strand of literature (e.g. Hyndman et al., 2012) assumes that subjects form beliefs based on historical data with specifically declining weights.

⁴ The primary objective of their analysis as well as ours is to understand qualitatively how the players model the behavior of others and respond to them, rather than to fit the data to a learning path toward taking certain actions per se. In this sense, our learning analysis departs from the existing literature on action-based learning, which as pointed out by Haruvy and Stahl (2012, p. 210) are “unhelpful” for analysis of rule learning across games, and most importantly does not allow the interpretation of how strategic sophistication develops with experience. Examples of action-based learning models include directional learning (Selten and Stoeker, 1986), reinforcement learning (Erev and Roth, 1998), and experience-weighted attraction learning (EWA) (Camerer and Ho, 1999), which considers both reinforcement and belief learning.

number of players who have already joined at the respective information set. Payoffs accrue when all players have moved. Player i 's eventual payoff from joining the club is denoted as $p_i(1, k)$, again k denotes the number of opponents that have joined in the respective node (i.e. when payoffs accrue, at the end of the game), and his payoff from not joining in response to k joining opponents is denoted as $p_i(0, k)$. The game is a club game if the payoff functions of all players satisfy these four restrictions.

$$\forall i \in N \forall k > l: \quad p_i(1, k) > p_i(1, l) \quad (1)$$

$$\forall i \in N \forall k < n: \quad p_i(1, k) - p_i(0, k) \neq 0 \quad (2)$$

$$\forall k < n \forall i < j: \quad p_i(1, k) - p_i(0, k) > p_j(1, k) - p_j(0, k) \quad (3)$$

$$\forall i \in N: \quad p_i(1, k) > p_i(0, k) \Leftrightarrow k \geq i - 1 \quad (4)$$

By Eq. (1), the payoffs from joining $p_i(1, k)$ are increasing in the club size, i.e. in k . By Eq. (2), payoffs are not degenerate. Eq. (3) ensures that players are labeled such that for all k , player 1 is most interested in joining the club, player 2 is second-most interested in doing so, and so on. Finally, Eq. (4) ensures that a club game satisfies *connectedness* in payoffs. Player 1 is generally best off joining the club, player 2 is best off joining *iff* at least one other player joins, player 3 is best off joining *iff* at least two other players join, and so on.

The remaining definitions of expected payoffs, sequential rationality, and Bayesian belief updating are standard, and a general understanding of these concepts is sufficient to understand the intuition underlying our theoretical results. Their formal definitions for club games are required only to prove the results formally. For this reason, these definitions as well as all proofs are relegated to [Appendix A](#).

3. Equilibrium analysis

First, we show that if players are sequentially rational and update beliefs by Bayes' Rule, all players join the club under all move sequences. The incompleteness of information and the fact that all players but one may prefer not having a club at all are thus outcome irrelevant.

Proposition 1. *Fix a strategy profile σ that is sequentially rational for some system of beliefs satisfying Bayes' Rule. Regardless of the move order, all players choose to "join" with probability 1 along the path of play.*

Proof. See [Appendix A](#). $\square$

To provide intuition, assume towards a contradiction that there would be a weak Perfect Bayesian equilibrium where some players do not join along the equilibrium path. Let $N' \subseteq N$ denote the set of these players. The remaining players $N'' = N \setminus N'$ join in all information sets on the equilibrium path, regardless of the move sequence. Now suppose that player i has to act in an information set (i, t, k) compatible with σ where at least one of the previously acting opponent did not join ($k < t - 1$). By Bayes Rule, player i 's belief is correct in that he does not assign positive probability to some $j \in N''$ not having joined the club, as it may have been some $j \in N'$. In any information set compatible with σ , the player to move therefore believes that all $j \in N''$ will still join if he sticks to σ . Now consider $i = \min N'$, i.e. the one with the lowest index under the assumed ordering Eq. (3), and an information set where j does not join. Since all move sequences have positive probability, all $i \in N''$ join along the path whether or not j joins, and by Eqs. (1) and (4) this implies that j is best off joining regardless of how the $i' \in N'$ decide—a contradiction. It follows that all players join along the path of play, i.e. in all information sets (i, t, k) ,

$$k = t - 1 \Rightarrow \sigma_i(t, k) = 1 \quad (\text{all join})$$

holds true. We will refer to these information sets (i, t, k) with $k = t - 1$ as those where *joining is iteratively dominant on path*. The remainder of this section investigates actions off the path. Let us begin with information sets where non-iterative dominance arguments suffice. In information sets where the requisite number of opponents have joined already, the respective players are best off joining, i.e. *joining is non-iteratively dominant*. In information sets where the requisite number of opponents cannot be reached anymore, they are best off not joining, i.e. *joining is non-iteratively dominated*.

$$k \geq i - 1 \Rightarrow \sigma_i(t, k) = 1 \quad (\text{dominant})$$

$$k + (n - t) < i - 1 \Rightarrow \sigma_i(t, k) = 0 \quad (\text{dominated})$$

In the remaining information sets (i, t, k) , for which we will show that *joining is iteratively dominant (off path)*, the following applies: (i) at least one player has chosen not to join, (ii) the number of players who have joined already does not suffice already, and (iii) the number of players who have decided not to join does not suffice to trigger "dominatedness" for i . In such information sets, i 's optimal action depends on whom he believes is (are) the player(s) who decided not to join. For example, assume that one player had decided not to join, one player is left to move after i , and i 's decision depends on

whether the subsequent player joins. If i believes that a move order applies according to which the subsequent player is the least enthusiastic player (with respect to joining the club, i.e. player n), then i should not join. For, in this case, the least enthusiastic player would not join regardless of i 's decision. However, if i believes that a move order applies according to which say player 1 moves last, then i is best off joining.

These information sets, where at least one opponent did not join, are off the path. Thus, they have a prior probability of zero and Bayesian updating does not restrict the beliefs in them. Player i may believe that the most enthusiastic player did not join while the least enthusiastic player is still to move and vice versa. Further, under perfection, if player 1 trembles toward “not join” with probability ε in all information sets on the path, player n trembles with probability ε^3 , and all other opponents tremble with probability ε^2 . It follows that as ε approaches 0, player i believes that player 1 did not join and player n is still to move. This holds true regardless of the prior probabilities of the various move orders.

Since player 1 is enthusiastic and n is not, the belief that 1 would be more likely to tremble than n is counter-intuitive, however. Such beliefs can be sustained under perfection only because relative tremble probabilities between players are unrestricted. More restrictive concepts such as properness restrict relative tremble probabilities between the various actions of individual players, and concepts such as strategic stability require robustness with respect to all tremble patterns—including counter-intuitive ones as those discussed above. Similarly, the global games approach to equilibrium refinement (Carlsson and van Damme, 1993; Frankel et al., 2003) and analyses of “robust equilibria” (Morris et al., 1995; Kajii and Morris, 1997; Ui, 2001) do not restrict relative payoff perturbations between players, and hence the implied restrictions on relative tremble probabilities are very weak.

Solution concepts derived from models of bounded rationality tend to induce the intuitive restrictions between players. For example, equilibrium refinement based on choice-theoretic models following e.g. McFadden (1976, 1984) generally restricts relative payoff perturbations between players, and hence we obtain falsifiable restrictions of relative tremble probabilities between players. Note that, while the original definition of quantal response equilibrium is not restricted in this manner (see McKelvey and Palfrey, 1995), most game-theoretic as well as choice-theoretic studies actually assume the multinomial logit specification implied by the extreme value distribution (see also Anderson et al., 1998), i.e. logit equilibrium. The following proposition establishes uniqueness of limiting logit equilibria in club games and characterizes the solution.

Proposition 2 (Logit equilibrium). *In the unique limiting logit equilibrium, player i joins in information set (i, t, k) if and only if this pays off when all $n - t$ subsequent players will choose to join (i.e. iff $k + (n - t) \geq i - 1$).*

Proof. See Appendix A. $\square$

This simple rule-of-thumb applies both on and off the equilibrium path in the limiting logit equilibrium. Implicitly, the players have the most optimistic beliefs about the order of moves: If some opponents did not join, then these must have been the least enthusiastic ones. For, the variances of payoff perturbations are constant between players and the payoffs of the least enthusiastic players require the least perturbation so that they do not join. As the variances tend to zero, the posterior probability of the assertion that it was the least enthusiastic players who did not join thus tends to 1.⁵

In logit equilibria, players deviate from equilibrium strategies due to payoff perturbations. Alternatively, they may deviate when they fail to anticipate their opponents' strategies correctly. This alternative idea underlies the “level- k ” model of reasoning.⁶ Level 1 players believe their opponents act non-strategically, level 2 players believe that their opponents believe they would respond to non-strategic players, and so on. This is related to assuming rationality, mutual knowledge of rationality, and so on, up to common knowledge of rationality at level $k = \infty$. The following result shows that extensive form rationalizability (EFR, see Pearce, 1984, and Battigalli, 1997) induces the same refinement effect as limiting logit equilibrium. That is, as level- k players “become more rational,” their beliefs approach the ones described above—without imposing any restrictions of the actions at level 0.⁷

Proposition 3 (Rationalizability). *The unique extensive form rationalizable strategy profile is the limiting logit equilibrium.*

⁵ Experimental evidence on the assertion that players that have less to lose are more likely to tremble is mixed, however (see e.g. Battalio et al., 2001, and Cooper and Van Huyck, 2003).

⁶ Our modeling approach largely follows the non-equilibrium models of Ho et al. (1998), Stahl (1998), Gneezy (2005), Costa-Gomes and Crawford (2006), and Crawford and Iriberri (2007).

⁷ Pearce defined EFR as an iterative elimination procedure, and in each iteration, the players' beliefs are restricted to those that are compatible with strategies that are not yet eliminated. In the first iteration, all players realize that player 1 generally joins, whereas players 2, ..., n may or may not join as first movers. As for the beliefs in the second iteration, this implies that in information sets where some player did not join in the first round, no player believes this would have been player 1 (since there are more “plausible” explanations in the sense that other explanations do not violate rationality). Iteratively applied, this induces the aforementioned refinement effect. Alternative refinements of rationalizability, e.g. perfect rationalizability (Bernheim, 1984) or proper rationalizability (Schuhmacher, 1999; Asheim, 2002), do not induce this refinement effect—and the reason is the same as it has been with the equilibrium concepts. Neither perfect nor proper rationalizability restricts tremble probabilities between players, although such restrictions are intuitive in club games: Less enthusiastic players are more likely to deviate. In turn, EFR and limiting logit equilibria induce such beliefs.

Proof. See [Appendix A](#). $\square$

In contrast, let us now discuss the respective predictions for the club game under *complete* information. The predicted outcome that all join is the same, but the actions off the equilibrium path differ. In contrast to above, the player now knows the identity of any player deviating from “all join”, which has two important implications. On the one hand, the equilibrium actions in information sets off the path will depend on the identity of the deviating player, i.e. strategies are a little more complex. On the other hand, and more importantly, players will be forced to abandon level- k beliefs for levels 2 and higher in many information sets off the path, e.g. when players deviate from non-iteratively dominant choices. Those deviations are identifiable for co-players under complete information, while incomplete information always allows to sustain the intuitive belief that one of the least enthusiastic players did not join, which in turn does not violate level 2 assumptions. Thus, the complete information club game does not allow players to maintain say level 2 beliefs throughout, while the incomplete information game does so for all “relevant” information sets.⁸ This implies that level 2 strategies are not identifiable under complete information, unless one simultaneously assumes high error rates; they are identifiable under incomplete information also for low error rates, besides being less complex.

4. Experimental design

4.1. Games and treatments

The payoff functions adopted in the experiment follow [Dixit \(2003\)](#) and are thus also special cases of the framework introduced above. A game has four players. Each has a “player type” indexed by $i = 1, 2, 3, 4$. The higher a player's index i , the lower is the player's benefit from joining the club. There are three payoff parameters. Parameter γ captures positive network externalities for those who join, τ captures negative externalities for those who do not, and β denotes a fixed benefit from joining. The payoffs from not joining, again denoted as $p_i(0, k)$ and joining, $p_i(1, k)$, in response to k joining co-participants are

$$p_i(0, k) = 20 \cdot [1 + i - \tau k] \quad \text{and} \quad p_i(1, k) = 20 \cdot [1 + \beta + \gamma k - i], \quad (5)$$

respectively. [Table 1](#) displays parameters and payoff functions in the experimental games.⁹ In relation to general framework analyzed above, $\gamma > 0$ implies that payoffs from joining are increasing in club size, Eq. (1), non-integer β yields non-degenerate payoffs Eq. (2), $dp_i(0, k)/di > 0$ and $dp_i(1, k)/di < 0$ yields the ordering of players by their incentives to join the club, Eq. (3), and $\gamma + \tau = 2$ yields connectedness Eq. (4) for all $\beta \in (2, 4)$.

We varied the strengths of incentives β and externalities γ, τ across treatments to obtain a 2×2 factorial design. Incentives in treatments B and D are stronger in absolute terms than in treatments A and C, while externalities in treatments C and D are stronger in absolute terms than in A and B. The manipulation of incentives affects QRE predictions without affecting level- k predictions (if the latter have high precision). The manipulation of externalities varies the proportion of players who would have been better off had the club not been started, which allows us to identify (and control for) social preferences if it affects subjects' choices.

Since all players join in equilibrium, $W = \sum_i p_i(1, 3)$ is the overall welfare.¹⁰ Thus, the club raises welfare if $\sum_i p_i(1, 3) > \sum_i p_i(0, 0)$, and it benefits player i if $p_i(1, 3) > p_i(0, 0)$. These two welfare relations vary systematically across the four treatments. In treatment A, the club reduces utility for three players out of four and it reduces the overall welfare. In treatments B and C, clubs are utility reducing for two players out of four, and it reduces the welfare in B, while it increases the welfare in C. In treatment D, clubs are utility improving for three players out of four and welfare improving overall.

This design allows for direct cross-comparability, to test the effects of each exogenously determined variable and to consider the robustness of the theory across different parameter sets. Most importantly, while controlling for social preferences, it serves to distinguish between level- k and QRE, as shown next.

4.2. Information sets, parameters and predictions

We define “behavioral types” by the reasoning mode a subject uses; this is not to be confused with the “player types” defined by the player's index in the club game. The club game separates level- k and QRE types mainly because it is *iteratively*

⁸ Exceptions are information sets where more than two players did not join (which is very rare in the four-player games analyzed below). For example, when a player $i > 1$ is to move last and learns that all co-players chose not to join, he would have to abandon say level 2 beliefs even under incomplete information though this would not affect his decision under connectedness.

⁹ The games were parameterized to obtain integer valuations to simplify the presentation of the games to the subjects. As above, there is complete information on the payoff structure and type distribution, but players do not know the order of moves.

¹⁰ This is, of course, a simplification as social welfare analysis largely depends on the specific application of the club game. On the one hand, summing individual payoffs to measure social surplus may make sense in an analysis where social welfare is well represented by those of the group agents. On the other hand, there is no reason why, a priori, and to refer to one of the original inspirations for the game as referred to in an earlier footnote, the social value of the activity should not take into account the potential negative externality for society from the activity, e.g. rioting. That being said, it is useful to see how far we can get if we assume that social welfare is purely based on the payoffs of the players involved.

Table 1

Experimental games with payoffs of the various player types in response to 0...3 co-participants joining the club.

Treatment A ($\beta = 2.1, \gamma = 0.5, \tau = 1.5$)						Treatment B ($\beta = 3.1, \gamma = 0.5, \tau = 1.5$)					
	"Join"	0	1	2	3		"Join"	0	1	2	3
Type 1	Yes	42	52	62	72	Type 1	Yes	62	72	82	92
	No	40	10	-20	-50		No	40	10	-20	-50
Type 2	Yes	22	32	42	52	Type 2	Yes	42	52	62	72
	No	60	30	0	-30		No	60	30	0	-30
Type 3	Yes	2	12	22	32	Type 3	Yes	22	32	42	52
	No	80	50	20	-10		No	80	50	20	-10
Type 4	Yes	-18	-8	2	12	Type 4	Yes	2	12	22	32
	No	100	70	40	10		No	100	70	40	10
Treatment C ($\beta = 2.1, \gamma = 1, \tau = 1$)						Treatment D ($\beta = 3.1, \gamma = 1, \tau = 1$)					
	"Join"	0	1	2	3		"Join"	0	1	2	3
Type 1	Yes	42	62	82	102	Type 1	Yes	62	82	102	122
	No	40	20	0	-20		No	40	20	0	-20
Type 2	Yes	22	42	62	82	Type 2	Yes	42	62	82	102
	No	60	40	20	0		No	60	40	20	0
Type 3	Yes	2	22	42	62	Type 3	Yes	22	42	62	82
	No	80	60	40	20		No	80	60	40	20
Type 4	Yes	-18	2	22	42	Type 4	Yes	2	22	42	62
	No	100	80	60	40		No	100	80	60	40

Fig. 1. Probability of joining for QRE and level- k players, depending on precision λ .

dominance solvable, which implies non-monotonic convergence of the predicted strategies as precision increases. Fig. 1, which plots the predicted choices of levels 0, 1, 2, and QRE types, all with logistic errors, illustrates this for two information sets. As discussed at the end of Section 3, the incompleteness of information allows to identify level- k strategies in the first place.¹¹ The figure distinguishes the most basic behavioral types: Level 0 randomizes uniformly ($\lambda_0 = 0$), level 1 logit responds ($\lambda_1 \geq 0$) to level 0, level 2 logit responds ($\lambda_2 \geq 0$) to level 1, and the equilibrium types stick to the logit equilibrium

¹¹ We use these information sets as examples for our illustration as they are observed particularly frequently in the experiment. Further differences on the levels of reasoning and precision required to find joining optimal exist in other information sets, but the basic qualitative aspects underlying type separation

(QRE) corresponding with some $\lambda_e \geq 0$. Due to the non-monotonicity, the strategies of the various types spread out over the full strategy space, ranging from 0% to 100% in many information sets, rather than converging from the low-precision strategy of joining with 50% directly to either 0% or 100%, which applies to games that are non-iteratively dominance solvable in all information sets. Information sets with non-iteratively dominant actions allow us to estimate the subjects' precision, and controlling for these, while iterative solvability allows us to identify the levels of reasoning.

In information sets where choices are non-iteratively dominance solvable, where *joining is non-iteratively dominant* and *joining is non-iteratively dominated* as discussed in Section 3, joining and not joining, respectively, is optimal for the player regardless of his conjecture concerning the remaining players. This applies in information sets (i, t, k) where $k \geq i - 1$ or $k + (n - t) < i - 1$ (see above). These information sets allow us to test for individual rationality when iterated arguments and conjectures about opponents are not required.

Hypothesis 4 (*Non-iterative dominance; $k \geq i - 1$ or $k + (n - t) < i - 1$*). Subjects join the club when joining is non-iteratively dominant, and do not join the club when joining is non-iteratively dominated.

In the remaining information sets, *joining is iteratively dominant* ($k < i - 1$), and play is either “on path” ($k = t - 1$) or “off path” ($k < t - 1$). Due to the iterations required, these information sets allow us to test for iterated reasoning and confidence in the rationality of co-players. As shown in the proof of Proposition 3, found in Appendix A, the structure of reasoning modeled in EFR implies that player type 2 (3) [4] requires two (three) [four] levels of reasoning for joining to be optimal, as an implication of the “iff” in the definition of connectedness. Requirements on iterative reasoning also decrease with one's position in the order of moves. It follows that as the level of reasoning required for joining to be individually optimal increases (decreases), in the face of bounded rationality we should then observe joining less (more) frequently.

Hypothesis 5 (*Iterative dominance; $k = t - 1$ or $k < t - 1$ and $k < i - 1$*). Whether or not subjects join the club when joining is iteratively dominant depends negatively on player type and positively on one's position in the order of moves.

Levels of reasoning can thus be identified if observed choices are sensitive to player type, as precision is identified by choices made in information sets that are non-iteratively dominance solvable and by varying the incentive β . Lower player types may be best off joining already with moderate precision, e.g. player 2 in treatment A. The existence of iteratively dominant information sets both on the path and off the path is important, as different levels of reasoning and precision are required for joining to be optimal. For example, Fig. 1 shows that the optimal choices of all other players depend on both their beliefs and the incentive parameter. In treatments A and C, players have comparably low incentives to join, and in B and D, the incentives are high. If incentives are low, then players join only if both level of reasoning and precision are high. For example, Fig. 1b shows that if player 3 joins in information set $(2, 0)$ in both treatments A and B, then he is identified as equilibrium type. A subject who joins in neither of these treatments is identified as level 1. If he joins in B but not in A, then he is level 2, and otherwise level 0. The same applies for player 4 in information set $(1, 0)$, see Fig. 1a. Note in particular how level 2 and equilibrium predictions diverge in these cases in treatment A. Thus, observed sensitivity to move positions, similarly to sensitivity to player type or to the incentive effects due to varying β , further restricts the precision parameters $\lambda_{0,1,2,e}$ for given levels of reasoning.

The rationalizability of an action is not sensitive to a player's position in the order of moves unless that position affects the information set's classification. This holds similarly—though not exactly—for level- k choices. Logit responses, in turn, are sensitive to variations in the expected payoffs, and comparing the joining rates across treatments with different incentives therefore show to which degree choices are affected by logistic errors. If incentives are high, for example, either high level or high precision suffices. Fig. 1 shows that generally more precision is required for a player with a level of reasoning greater than 0 to join in treatment A than in B, because there are less incentives to join the club in treatment A than in B. Equilibrium types join in treatment A only if they play a high precision equilibrium as they need be convinced that sufficiently many opponents will join. In turn, this requires high precision on their side. Thus, we can distinguish low-precision equilibrium players, namely those who join in B but not A, and high-precision equilibrium players namely those who join in both B and A.

Hypothesis 6 (*Required level of reasoning and incentive effects*). The probability of joining the club is negatively related to the required level of reasoning, and positively related to the relative gains from joining the club.

It is by manipulating treatments of the club game in this way that we can achieve the separation of types. By cross-checking behavior across types 1 to 4, treatments A to D, and all 32 observations per subject, we are thus able to sharply distinguish different behavioral types, namely QRE, level- k , and level 0 types. Indeed, the posterior probabilities estimated below concur. While all players join the club if they are rational and update beliefs by Bayes' Rule, their behavioral pattern is

are similar. Our computer scripts to compute level- k and QRE predictions are available on request (QREs are computed using Gambit, see McKelvey et al., 2007).

diverse if we consider bounded levels of reasoning, i.e. level- k for small k , or bounded precision in payoff maximization, i.e. logit equilibrium for finite λ . Based on the results of existing studies on normal form games, see e.g. Crawford and Iriberri (2007) and the related literature cited therein, we also expect to observe a mixture of behavioral types in our extensive form game.

An alternative assumption to one where subjects apply the same belief rules even with experience, i.e. behavioral types are constant over time, is that subjects apply different rules with experience. To analyze reasoning modes at the individual level, we kept the combination of information set types that a subject can encounter during play specific to his player type and constant across treatments, thus allowing us to distinguish level- k reasoning from quantal response by analyzing observed play over time and across treatments. This, in turn, will reveal whether they individually change their belief rules over time, i.e. whether subjects approach equilibrium via the level- k path from levels 0, 1, 2, to equilibrium, or the QRE path from level 0 via increasingly precise QREs to equilibrium.

Hypothesis 7 (*Belief rule changes*). The sample is heterogeneous and contains a mixture of level- k and QRE types. With experience, subjects adapt their belief rules via the level- k path or, alternatively, via the QRE path.

Repeated observations of experimental play allow us to test if subjects converge to a steady state in which the proportions of behavioral types in the sample stay constant. In Section 6, we propose a Markov-switching rule learning model developed to analyze learning patterns and convergence, and also test if this outperforms a reinforcement rule learning model. The three main implications of this test are: 1) to ascertain that learning takes place, and if so; 2) to inform us of the strategic learning outcome in the context of equilibrium refinement in our extensive form game, and; 3) the process by which such learning outcomes are achieved. As discussed in the Introduction, the learning process can be one where subjects randomly experiment on alternative decision rules reinforced by experienced success as modeled by Haruvy and Stahl (2012), or by a more systematic process of transitions between rules identifiable by our Markov-switching model.

Hypothesis 8 (*Convergence*). With experience, the proportions of types in the sample converge to a steady state.

4.3. Procedure and logistics

The experiment was conducted in the experimental economics laboratory at the European University Viadrina, Frankfurt (Oder), Germany. The experimental instructions and questionnaire are found in Appendix C. Subjects were recruited from an email list consisting of students from the faculties of Cultural Science, Business and Economics, and Law. We conducted six sessions with twelve subjects each. Each session consisted of four stages,¹² each of eight rounds. Each stage involved different treatments. We presented treatments either in the sequence A-C-B-D or in the inverse sequence D-B-C-A to control for order effects. Subjects were randomly allocated a type at the start of the experiment; a subject's type was maintained throughout the experiment.

Each round consisted of a game where subjects were matched into groups of four and asked in sequence if they wanted to "become a member of a club," and to choose *yes* or *no*. This phrasing was chosen as being intuitive and simple to understand. Each group contained player types 1, 2, 3, and 4 (corresponding to preference index $i = 1, 2, 3, 4$ labeled as P, Q, R, and S in the experiment), and this was known to the subjects. The payoff tables for all four players in a group were shown on the computer display. For each round, subjects were randomly allocated to a position in the order of moves. Subjects were informed of how many had moved, and how many had joined the club (chosen 'yes'), before moving. At the end of each round, subjects were given feedback of the total number of subjects in their group who had joined, and their own earning. In each new round, subjects were re-matched into new groups of four. This allows for experience, and eliminates reputation and supgame effects.

At the beginning of the experiment, subjects were randomly allocated to their seats. Then, they were asked to read the experimental instructions, provided on printed sheets, and to answer a short control questionnaire for us to check their understanding. Subjects in doubt were verbally advised by the experimental assistants before being allowed to begin. Each computer terminal was partitioned, so that subjects were unable to communicate via audio or visual signals, or to look at other computer screens. Decisions were thus made in privacy. At the end of the experiment, subjects were informed of their payments, and asked to privately choose a code-name and password. This was used to anonymously collect their payments from a third party one week after the experiment. Each subject was given a €6 participation fee and the earnings from two randomly chosen "winning rounds" from stages 1–4, and one from stage 5. Each experimental point was worth €0.10. The average pay-out per subject was €20.83 for approximately 1.5–2 hours per session (which for example is more than the local wage of €8 per hour for research assistants).¹³

¹² We also had an additional brief stage to try to measure level of reasoning by the subjects using a different task (a 'p-beauty contest' game), as in principle finding the rationalizable equilibrium in the club game may depend on different levels of reasoning for different types. However, this measure was not related to behavior in the earlier part of the experiment and it will not be referred to further.

¹³ Our focus on reasoning modes and rule learning dynamics requires a large number of observations per subject on various information sets and treatments, implying longer sessions, rather than having a larger number of subjects but with shorter experimental duration. As shown in later sections and in the supplementary material, we are able to achieve robust results with our sample size.

Table 2

Relative frequency of “Join” (and “Not join” in the last row).

Treatment	Overall				First half of experiment				Second half of experiment			
	A	B	C	D	A	B	C	D	A	B	C	D
Type 1	0.99	0.99	1	0.99	0.99	1	1	0.97	1	0.99	1	1
Type 2	0.89	0.94	0.91	0.97	0.82	0.99	0.88	0.97	0.96	0.9	0.94	0.96
Type 3	0.63	0.89	0.74	0.92	0.49	0.92	0.61	0.88	0.78	0.86	0.88	0.97
Type 4	0.26	0.69	0.34	0.74	0.14	0.76	0.25	0.69	0.38	0.61	0.43	0.79
Aggregate across types												
Join	0.69	0.88	0.75	0.9	0.61	0.92	0.68	0.88	0.78	0.84	0.81	0.93
Not join	0.31	0.12	0.25	0.1	0.39	0.08	0.32	0.12	0.32	0.16	0.09	0.07

Table 3

Frequency of rationalizable actions.

Treatment	Overall				First half of experiment				Second half of experiment			
	A	B	C	D	A	B	C	D	A	B	C	D
Type 1	0.99	0.99	1.00	0.99	0.99	1	1	0.97	1	0.99	1	1
Type 2	0.89	0.94	0.91	0.97	0.82	0.99	0.88	0.97	0.96	0.90	0.94	0.96
Type 3	0.67	0.90	0.76	0.94	0.56	0.92	0.63	0.90	0.79	0.88	0.90	0.97
Type 4	0.44	0.71	0.49	0.79	0.39	0.79	0.46	0.74	0.50	0.63	0.51	0.85

Table 4

Social surplus (measured as sum of payoffs).

Treatment	All join	None join	Actual joining
A	168	280	147
B	248	280	224.67
C	288	280	239.18
D	368	280	339.18

5. Overview of observed behavior

In this section, we describe the main patterns of behavior in the experiment and relate them qualitatively to the models discussed above, logit equilibrium, level- k and rationalizability. (Table 2 for relative frequencies of joining for each type and treatment in the first and second halves of the experiment and overall.) Overall 84% of all actions in the experiment were rationalizable according to EFR, and rationalizable actions were more frequent in the second half (86.1%) of the experiment than in the first half (81.2%). Thus, EFR predicts fairly well and better with experience. Table 3 shows the relative frequencies of rationalizable actions for all player types and all treatments, both overall and for the first and second halves of the experiment separately. Table 5 further splits joining rates down for each of the information sets. Most deviations from the predictions concern decisions where subjects who “should have” joined did not join. Summarizing Table 5, for treatments A, B, C and D respectively, of the decisions not to join, in aggregate 20%, 13%, 18%, and 18% were rationalizable, for types 3 11%, 6%, 8% and 18% were rationalizable, while for types 4 28%, 18%, 24% and 22% were rationalizable.

Individual differences in welfare across types naturally emerge due to heterogeneous preferences for the club, as shown in Table 1. We consider social surplus to be the sum of payoffs, as defined in Section 4. Table 4 shows the social surplus of the hypothetical cases when all or none joined the club, the actual social surplus observed in our experiment, whether a club is socially desirable, i.e. the social surplus is less when everybody joins the club instead of when nobody joins the club, and a comparison of the observed outcome with those of the hypothetical cases. In our experiment, the observed social surplus was less than if all had joined the club. With the exception of treatment D, the observed social surplus was also less than if none had joined the club.

Table 5 reports the joining rates per information set. The overall pattern is that (i) keeping (t, k) fixed, the higher the player type i , the smaller the joining rate; (ii) keeping type i and number of players that did not join $(t - 1 - k)$ fixed, the earlier the mover position t , the smaller the joining rate, and; (iii) keeping types i and mover positions t fixed, the higher the number of players still required to join in order to trigger optimality of one's joining $(i - 1 - k)$, the smaller the joining rate. All three of these qualitative observations point toward the same direction: the less the incentive to join the club, and the more uncertain it is that the requisite number of joining co-participants would be reached, the lower the joining rate. Fig. 1 shows that the overall pattern of deviations from the equilibrium predictions is intuitive and seems to match qualitatively with logistic errors and finite-level reasoning. This is analyzed in further detail next.

Table 6a looks at information sets with dominant actions, i.e. iterative arguments are not required. In 97% of the cases, subjects join when joining is non-iteratively dominant, and they do not join 87% of the time when joining is non-iteratively dominated. Overall, the dominant action is chosen in 96% of the cases, which is evidence for Hypothesis 5.

Table 5Joining rates (numbers of observations) for all information sets (i, t, k) .

Treatment A					Treatment B				
Type $i = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.97(35)	1(16)	1(15)	1(5)	$k = 0$	0.97(31)	1(9)	1(3)	1(2)
$k = 1$		1(16)	1(22)	1(10)	$k = 1$		1(22)	1(16)	1(2)
$k = 2$			1(8)	1(12)	$k = 2$			1(21)	1(11)
$k = 3$				1(5)	$k = 3$				1(27)
Type $i = 2$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 2$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.84(31)	0.71(14)	0.43(7)	(0)	$k = 0$	0.92(37)	1(10)	0.33(3)	(0)
$k = 1$		1(23)	1(19)	0.95(19)	$k = 1$		1(33)	1(11)	0.75(4)
$k = 2$			0.9(10)	1(18)	$k = 2$			0.96(25)	1(9)
$k = 3$				0.67(3)	$k = 3$				0.92(12)
Type $i = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.59(39)	0.27(15)	0(2)	0(1)	$k = 0$	0.86(43)	0.6(10)	(0)	(0)
$k = 1$		0.56(27)	0.42(12)	0(3)	$k = 1$		0.83(24)	0.89(9)	1(1)
$k = 2$			1(14)	0.96(24)	$k = 2$			1(24)	1(13)
$k = 3$				1(7)	$k = 3$				1(20)
Type $i = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.15(39)	0.1(10)	(0)	(0)	$k = 0$	0.36(33)	0(2)	(0)	(0)
$k = 1$		0.13(23)	0.1(10)	0.5(2)	$k = 1$		0.71(34)	0.33(3)	(0)
$k = 2$			0.32(25)	0(11)	$k = 2$			0.86(29)	0.5(8)
$k = 3$				0.71(24)	$k = 3$				0.94(35)
Treatment C					Treatment D				
Type $i = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 1$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	1(28)	1(15)	1(8)	1(6)	$k = 0$	1(40)	1(7)	(0)	0.5(2)
$k = 1$		1(24)	1(14)	1(4)	$k = 1$		1(27)	0.9(10)	1(2)
$k = 2$			1(17)	1(18)	$k = 2$			1(29)	1(9)
$k = 3$				1(10)	$k = 3$				1(18)
Type $i = 2$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 2$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.93(45)	0.83(12)	0(4)	(0)	$k = 0$	0.97(37)	0.75(8)	(0)	(0)
$k = 1$		0.95(19)	1(22)	0.92(12)	$k = 1$		1(30)	1(12)	1(1)
$k = 2$			1(13)	0.92(12)	$k = 2$			0.96(23)	1(4)
$k = 3$				0.8(5)	$k = 3$				0.97(29)
Type $i = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 3$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.62(45)	0.5(10)	0(2)	(0)	$k = 0$	0.87(30)	1(4)	0(2)	(0)
$k = 1$		0.87(31)	0.5(12)	0(1)	$k = 1$		0.88(32)	1(6)	(0)
$k = 2$			1(13)	0.91(23)	$k = 2$			0.96(28)	1(13)
$k = 3$				1(7)	$k = 3$				1(29)
Type $i = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	Type $i = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
$k = 0$	0.08(26)	0(7)	(0)	(0)	$k = 0$	0.62(37)	(0)	(0)	(0)
$k = 1$		0.31(26)	0.2(5)	0(1)	$k = 1$		0.75(36)	0(2)	(0)
$k = 2$			0.35(34)	0.08(12)	$k = 2$			0.88(32)	0.14(7)
$k = 3$				0.76(33)	$k = 3$				0.93(30)

Note:

 (i, t, k) information set of player i moving in position t when k co-participations have joined;

(red) joining is not rationalizable;

(yellow) joining is uniquely EF-rationalizable, but not uniquely trembling-hand perfect;

(white) joining is the dominant action, without iterative eliminations.

Table 6

Aggregate joining rates in various types of information sets.

(a) Joining is non-iteratively dominant/dominated						
(i, t, k)	joining is ...	$i = 1$	$i = 2$	$i = 3$	$i = 4$	All types i
$k \geq i - 1$	dominant	573/576	356/368	211/215	103/122	1243/1281 = 0.97
$k + (n - t) < i - 1$	dominated	n.a.	0/0	1/12	11/80	12/92 = 0.13

(b) Joining is iteratively dominant on path						
(i, t, k)	(2, 1, 0)	(3, 1, 0)	(3, 2, 1)	(4, 1, 0)	(4, 2, 1)	(4, 3, 2)
join	138/150	114/157	90/114	43/135	62/119	73/120

Aggregate Type $i = 1$: 138/150 = 0.92. Type $i = 2$: 204/271 = 0.75. Type $i = 3$: 178/374 = 0.48.

(c) Joining is iteratively dominant off path		
(2, 2, 0)	(3, 2, 0)	(3, 3, 1)
36/44	19/39	25/39

(d) Joining rates when it is iteratively dominant on versus off path		
(2, 1, 0) vs. (2, 2, 0)	(3, 1, 0) vs. (3, 2, 0)	(3, 2, 1) vs. (3, 3, 1)
$\frac{138}{150} \approx \frac{36}{44}$	$\frac{114}{157} > \frac{19}{39}$	$\frac{90}{114} > \frac{25}{39}$

Note: “ $\approx$ ” indicates insignificance at $\alpha = 5\%$ in Wilcoxon paired-sample tests using session-level data, “ $>$ ” indicates significance.**Table 7**

Significance of externality and incentive effects across treatments.

Player type	$i = 1$	$i = 2$	$i = 3$	$i = 4$
Incentive (A + C vs. B + D)	$\frac{287}{288} \approx \frac{285}{288}$	$\frac{259}{288} < \frac{275}{288}$	$\frac{198}{288} < \frac{261}{288}$	$\frac{86}{288} < \frac{206}{288}$
Externality (A + B vs. C + D)	$\frac{286}{288} \approx \frac{286}{288}$	$\frac{264}{288} \approx \frac{270}{288}$	$\frac{219}{288} \approx \frac{240}{288}$	$\frac{136}{288} \approx \frac{156}{288}$

Note: “ $\approx$ ” indicates insignificance at $\alpha = 5\%$ in Wilcoxon paired-sample tests using session-level data, “ $>$ ” indicates significance.

Result 9 (Non-iterative dominance). Subjects act as predicted 97% (87%) of the time in information sets where joining is non-iteratively dominant (dominated).

Table 6b looks at decisions in information sets on the path where joining is iteratively dominant. In total, subjects join 65% of the time. The discrepancy from equilibrium predictions under EFR and LLE are decreasing in player type, with 92%, 75%, 48% for types 2, 3, 4, respectively. Holding the player type constant, they are also increasing in one's position in the order of moves. These patterns have been hypothesized above and are compatible with logistic errors; the former one also with iterative reasoning.

Table 6c looks at information sets off the path. Again, joining is iteratively dominant, and in total, subjects join 66% of the time. This suggests that the distinction between information sets “on path” and “off path” is on average not necessary. Table 6d shows, however, that joining differs in “adjacent” information sets, i.e. in information sets differing only in the move position t of the player, depending on whether they are on or off the path. As discussed above, this, too, is qualitatively compatible with logistic errors.

Result 10 (Iterative dominance). Subjects join the club 65% (66%) of the time in information sets on (off) the equilibrium path where it is iteratively dominant to join. The average joining rate decreases with player type and increases with one's position in the order of moves.

The assumption of logistic errors is also compatible with the sensitivity of the joining rates to the treatment parameters. In particular, joining rates should not be independent of the strength of incentives of joining β if subjects play logit responses. Since this affects both one's incentives and the opponents' incentives, the effect should reinforce itself and hence it should be significant. Table 7 shows that as predicted joining rates are higher in treatments with stronger incentives for player types 2 to 4.¹⁴ In turn, the effect of varying the strength of externalities γ , τ is insignificant for all player types, as is predicted for players with logit responses and utility functions based solely on pecuniary payoffs. This does not rule out the relevance of social preferences altogether, but it suggests that social preferences are not of primary relevance in our games, which we analyze more specifically below.

¹⁴ This prediction is supported at the 5% level in Wilcoxon paired-sample tests using session-level data, as the p -value is about 0.018 in all cases.

Table 8

Results of probit regression with random effects (subject and session).

Intercept	DOMINANT	DOMINATED	UNCLEAR	LEVREAS	INCENTIVE	EXTERNALITY	TYPE
2.257** (6.77)	−0.038 (−0.16)	−2.880** (−9.11)	−0.607** (−4.35)	−0.651** (−6.69)	0.870** (10.64)	0.174* (2.28)	–
2.646** (6.93)	0.320 (1.01)	−2.023** (−4.538)	−0.573** (−3.099)	−0.352* (−2.511)	0.855** (7.839)	0.179 (1.737)	−0.459** (−4.426)

Note: * and ** denote significance at 0.05 and 0.01 level, respectively; t -values are given in parentheses.

DOMINANT = 1 when joining is dominant, DOMINATED = 1 when joining is dominated, UNCLEAR = 1 in information sets off the path where joining is iteratively dominant, LEVREAS = level of reasoning required in this information set (according to EFR), INCENTIVE = 1 in treatments B and D, EXTERNALITY = 1 in treatments C and D, TYPE is the player index i .

Finally, in order to confirm the qualitative compatibility of joining with logistic errors and finite-level reasoning, we conducted multivariate regression analysis of the joining rates. The model is multinomial probit with nested random effects at the levels of subjects nested in sessions. Table 8 reports the results. This analysis shows that the required level of reasoning is indeed significant, being in information sets off the path costs about one level of reasoning, while the effect of increasing incentives corresponds with that of 1–2 levels of reasoning. The former points toward level- k models, while the latter points toward logistic errors. Note that these results continue to hold even when we control for type i , although the player type correlates closely with both the required level of reasoning and the individual incentives to join.

Result 11 (Required level of reasoning and incentive effects). *The probability of joining is negatively related to the required level of reasoning, and this indicates the relevance of iterative reasoning. The probability of joining is positively related to the relative gains from joining the club, and this indicates the relevance of logit response.*

These results support Hypotheses 4–6. While behavior is much in line with that predicted by LLE and EFR, the observed deviations from equilibrium are compatible with weakening the assumption of best response toward logit response and with weakening that of strategic equilibrium toward level- k reasoning. This supports the approach taken in the next section, where we shall test Hypothesis 7 on type heterogeneity and belief rule changes, and Hypothesis 8 on convergence to a steady state.

6. Analysis of belief rules and rule learning

6.1. Basic specification

We hypothesize that subjects are heterogeneous with respect to their level of reasoning. The standard econometric approach to capture this kind of heterogeneity is the static mixture model (McLachlan and Peel, 2000), which we adopt. We also adopt the standard choices with respect to the error distribution (logistic errors with precision λ) and level 0 behavior (uniform randomization). In the model, C denotes the finite set of model components one for each reasoning mode, which defines behavioral types e.g. $C = \{0, 1, 2, e\}$; $\lambda = (\lambda_c)_{c \in C}$ denotes the respective precision parameters and $(\rho_c)_{c \in C}$ with $\sum_c \rho_c = 1$ denotes the component weights. Given λ , the choice probabilities of subjects in the various components are computed straightforwardly (logit equilibria using Gambit, see McKelvey et al., 2007). Using $o_{s,t}$ to denote the decision of subject $s \in S = \{1, \dots, 72\}$ in round $t \in T = \{1, \dots, 32\}$ and $\Pr(o_{s,t}|c, \lambda)$ as the theoretical probability of $o_{s,t}$ (conditional on s being of type c with precision λ), the likelihood of (λ, ρ) given data set $O = (o_{s,t})$ is

$$L(\lambda, \rho|O) = \prod_{s \in S} \sum_{c \in C} \rho_c \prod_{t \in T} \Pr(o_{s,t}|c, \lambda). \quad (6)$$

The model parameters are estimated by maximizing the likelihood function by an EM algorithm (Bilmes, 1998) and standard errors are obtained by bootstrapping.¹⁵

We will consider two different model specifications.¹⁶ The first contains levels 0–2 and one QRE component. The QRE component will be estimated to be a high precision one, i.e. it contains subjects that play the equilibrium prediction with high precision. The second specification contains level 0 and three QRE components estimated to be of low, medium, and

¹⁵ A reviewer suggested to use EM and forward–backward algorithms in order to be able to bootstrap standard errors, which is greatly appreciated and helped in particular with the Markov-switching model discussed below. Our original approach used the derivative-free NEWUOA algorithm of Powell (2008) first and then a Newton–Raphson algorithm to verify convergence, as well as many starting values, to maximize the likelihood jointly over all parameters. With this approach we also found the global maximum, but it would have been too time-consuming to bootstrap standard errors. The maximum likelihood estimates are the same for both approaches, and the bootstrapped standard errors are very similar to the asymptotic ones derived from the information matrix, but only bootstrapping allowed us to obtain standard errors for derived statistics such as transition probabilities in the Markov-switching model. See Appendix B for more details.

¹⁶ The supplementary material contains an analysis of alternative specifications, showing that further components are not required (such as models containing level 3 choices or further QRE components) and also that models with less components do not suffice.

high precision types (the number of components is equal across both models for comparability). In the rule learning models discussed below, their distinction allows us to observe whether subjects approach equilibrium behavior via the “level- k path” transiting through levels 1 and 2, or via the “QRE path” transiting through intermediate precisions of logit equilibrium.

In addition, we control for social preferences by allowing that subjects feel spiteful toward opponents earning more than them. The relevance of social preferences turns out to be small in our analysis, and none of our key results depend on allowing for this. The utility of $i \in N$ is $u_i = p_i - \sum_{j: p_i < p_j} \alpha * p_j$ for all $\mathbf{p} \in \mathbb{R}^N$.¹⁷ Of course, social preferences may matter more in other contexts or even in club games with stronger externalities, for example, and this can be examined in future research.

6.2. Models of rule learning

It is intuitive to assume that the mode of reasoning would adapt as subjects gain experience, as subjects can infer their opponents’ level of reasoning from their actions. The standard assumption in experimental analyses, however, is that the mode of reasoning is constant across an experimental session. Two exceptions are Stahl (1996) and Stahl (2000). In this section, we relax this standard assumption and examine if and how they adapt.

To this end, we partition the set of experimental games into segments and generalize the finite mixture model by allowing that subjects have a segment-dependent reasoning mode $c(q)$ —rather than a constant mode $c \in C$ as in static mixture models. We distinguish two different segmentations of the 32 experimental games, either into 4 segments of eight games each or into 8 segments of four games each. The former approach partitions the experiment precisely into its four stages, namely rounds $1 \dots 8, 9 \dots 16, 17 \dots 24, 25 \dots 32$, and thus allows that subjects update their beliefs when a new stage is entered. The second approach partitions it into the four stages and each stage also into its first half and second half, namely rounds $1 \dots 4, 5 \dots 8, 9 \dots 12, \dots$

In the following, let Q denote the set of segments of the experiment (e.g. the set of stages), and let c be aforementioned function mapping segments $q \in Q$ to reasoning modes $c(q) \in C$. The set of such functions is C^Q . Thus, each function $c \in C^Q$ describes a learning path $c(1), c(2), c(3), \dots$ across segments in the level- k /QRE space. Let $\Pr(c|\rho, \omega)$ denote the probability of path c given model parameters ρ, ω as discussed shortly, and let $T(q) \subset \{1, \dots, 32\}$ denote the subset of experimental rounds constituting segment $q \in Q$. The static mixture model imposes the restriction $c(q+1) = c(q)$ for all q , and that non-constant paths have probability zero, while generalized models allow for alternative paths. The correspondingly generalized likelihood function is

$$L(\lambda, \rho, \omega|O) = \prod_{s \in S} \sum_{c \in C^Q} \Pr(c|\rho, \omega) \prod_{q \in Q} \prod_{t \in T(q)} \Pr(o_{s,t}|c(q), \lambda). \quad (7)$$

Our analysis contrasts two approaches to model paths c and their probabilities $\Pr(c|\rho, \omega)$.

Markov-switching rule learning. Markov-switching models are widely used to model time-dependent regime changes and they are immediate generalization of static mixture models (see e.g. Krolzig, 1997; Frühwirth-Schnatter, 2006). The defining idea is the Markov assumption that the probability of being in mode $c(q+1)$ in the next segment depends only on one’s current mode of reasoning $c(q)$. The model may express the idea that subjects advance systematically through the various modes of reasoning, say from level 1 via level 2 to equilibrium, each step taken with a certain probability after a segment. The probability of moving from mode c' to mode c'' after the current segment is denoted as $\omega(c', c'')$, with $\sum_{c'' \in C} \omega(c', c'') = 1$ for all $c' \in C$. For example, the transition probabilities if the current mode is level 1 may be such that the subject stays with probability 70% in level 1, advances with 30% to level 2, and moves to either level 0 or equilibrium with probability zero. The probability of being in mode $c' \in C$ in the first segment is $\rho_{c'}$, as above. Overall, the probability of path $c \in C^Q$ is

$$\Pr(c|\rho, \omega) = \rho_{c(1)} \cdot \prod_{q \geq 1} \omega[c(q+1), c(q)], \quad (8)$$

and substituting this into the likelihood Eq. (7) yields the Markov-switching rule learning model. We estimate the parameters using the forward–backward algorithm (Bilmes, 1998).

Reinforcement-based rule learning. The alternative approach is one of reinforcement learning. The idea of reinforcement on actions is that subjects are experimenting by choosing actions randomly, while positive experiences increase the probability to use the respective action in the future, whereas negative experiences decrease the probability. Haruvy and Stahl (2012) applied this idea to “rule learning”. In their model, subjects experiment with different rules and reinforce on rules

¹⁷ Again, the supplementary material contains an analysis of alternative utility functions, including egoism, linear altruism, Fehr–Schmidt inequity aversion, but the simple form of conditional envy considered here fits best. Note that Fehr–Schmidt inequity aversion would be equivalent in our setting to strong reciprocity (e.g. if the subject who moves second observes his predecessor to join, he or she would consider it most likely that this was player 1 and refuse to join to punish player 1 since he or she has received a lower payoff). The supplementary material also considers time-varying altruism weights and further refinements, but none yielded significant improvements in relation to the Markov-switching models analyzed below.

rather than actions. In contrast to Markovian rule learning, reinforcement-based rule learning implies fairly stochastic learning paths. Subjects are assumed to experiment randomly with rules rather than adopting them progressively and perhaps monotonically.

Formally, the reinforcement model considered here follows Haruvy and Stahl (2012).¹⁸ Let $\psi(c, q)$ denote the log-propensity for using rule $c \in C$ in segment $q \in Q$. The probability of reasoning path c is the product (over all segments) of logistic propensities $\psi(c(q), q)$ for choosing reasoning mode $c(q)$ in segment q .

$$\Pr(c|\rho, \omega) = \prod_{q \in Q} \frac{\exp\{\psi(c(q), q)\}}{\sum_{c' \in C} \exp\{\psi(c'(q), q)\}} \quad (9)$$

The initial propensities $\psi(c, 1)$, $c \in C$, are random variables with distribution $\mathcal{N}(\rho_c, \sigma_c)$. After each segment q , the propensities are subject to reinforcement. In order to define the process, we use π_q to denote the realized payoffs in segment q ; those have weight ω_2 in the updating process, while ω_1 denotes inertia. Following Haruvy and Stahl (2012), we also allow for partial digression to the initial propensities if a new experimental stage is entered (where a different game is being played). Let the weight of the current propensities at stage transitions be ω_3 . Using the indicator $I_q \in \{0, 1\}$ to indicate a stage transition after segment q , propensities are updated as

$$\psi(c, q+1) = \frac{(\omega_1 \psi(c, q) + \omega_2 \pi_q) \omega_3 + I_q \psi(c, 1)(1 - \omega_3)}{\omega_3 + I_q(1 - \omega_3)}. \quad (10)$$

6.3. Results

Table 9 summarizes the main results. In addition to the static mixture model containing levels 0–2 and one equilibrium component, our analysis considers both Markov and reinforcement-based generalizations of this model, over either four or eight segments. Further, we allow for both, learning along the “level- k path” involving transitions across levels of reasoning, and learning along the “QRE path” involving transitions across level 0 and QREs with increasing precision. The supplementary material analyzes the robustness of our findings based on these models, showing that models allowing for alternative segmentation or segment-dependent precision/altruism do not improve significantly upon those models discussed here. The supplementary material also shows that the qualitative results do not change in these alternative specifications.

The quantitative model fit show in Table 9a indicates that both Markov switching rule learning models along the level- k path improve significantly on the static mixture model with constant belief rules, in support of Hypothesis 7. Neither the reinforcement rule learning model nor the Markov switching rule learning model along the QRE path does so. That is, our subjects do not seem to explore the various modes of reasoning via strategic experimentation and reinforcement, but via the more systematic Markov process, and they choose they level- k path rather than the QRE path, which confirms our previous qualitative observation that the required level of reasoning affects joining probabilities significantly. Further, as shown in Table 9b, the log-likelihood of the Markov model over eight segments is -606.11 , which is close to the hypothetical “maximum” $LL = -584.1$ resulting from the upper benchmark of predicting the relative frequencies in all treatments correctly. Thus, the Markov model indeed fits the observations.

The different results implied by the static mixture model and the dynamic Markov switching model are illustrated in Table 9b. As the large improvement in the BICs from 692.23 to 646.74 suggests, belief rules are changing dynamically rather than being static. Nonetheless, if we were to restrict ourselves to static belief rules, we would reach the conclusion that one belief rule, namely QRE with intermediate precision, fits the vast majority of subjects (79.5%) and that there are no subjects at level 1. In the static model, the QRE component happens to capture average behavior across segments rather well. However, the dynamic analysis reveals behavior is better explained by a model in which player beliefs vary over time. The relatively noisy “average component” fitted to the QRE behavioral type and the corresponding large population share masks these more complicated behavioral dynamics. Rather, the size of this QRE component appears to be an artefact of making the inadequate assumption that belief rules be constant. In the more adequate dynamic model, most subjects are level 1 or level 2 initially (49.2% and 41.1%, respectively), while only few play the QRE from the outset (6.8%).

The transition probabilities (Table 9b) underline the system behind the ensuing transitions across reasoning modes. First, no subject either enters or leaves level 0. The small set of level 0 subjects, i.e. the 3% of subjects randomizing uniformly, is constant. The remaining subjects flow across levels 1, 2, and QRE behavior. Specifically, level 1 subjects most likely (75.9%) upgrade toward level 2, as one would assume based on iterated dominance. Otherwise, they stay at level 1 (24.1%); none of them makes the whole jump toward equilibrium behavior immediately from level 1. Subjects at level 2, in turn, most likely stay there (57.6%), but they may also either learn QRE beliefs (29.4%) or downgrade to level 1 beliefs (13.0%). QRE subjects, finally, most likely maintain their beliefs (74.1%), but may also regress to level 1 (17.9%) or level 2 (7.9%). This pattern of beliefs is plausible. There is a learning pattern from level 1 to level 2, and from level 2 to QRE. When deviations relative to expectations are observed by a level 2 subject, or deviations relative to equilibrium play are observed by a QRE subject,

¹⁸ The only difference will be that we use the standard definition of level 0, uniform randomization, while they used a more complex construct that partially includes herding.

Table 9

Summary of the Markov transition analysis.

(a) LR tests on $BIC = -LL + \#Pars/2 \cdot \log \#Obs$ (less is better)

	#Pars	BIC	A	B	C	D	E	F
A Static mixture (L1,L2,Eq), constant rule	7	692.23		»»	»»	=	=	=
B Markov (L1,L2,Eq), four segments	19	675.45	««		»»	=	=	<
C Markov (L1,L2,Eq), eight segments	19	646.74	««	««		««	««	««
D Markov (QRE,QRE,Eq), eight segments	19	687.39	=	=	»»		=	=
E Reinforce (L1,L2,Eq), four segments	25	690.92	=	=	»»	=		««
F Reinforce (L1,L2,Eq), eight segments	25	716.94	=	>	»»	=	»»	

Note: The null $LL(\text{Row model}) = LL(\text{Column model})$ is evaluated in nested/non-nested Vuong tests using the BIC correction (as suggested by Vuong, 1989, Eq. 5.9). ««, «, < indicate significant rejections at 0.01, 0.05, 0.1 (resp.) in favor of the row model, and »»», », > do so in favor of the column model.

(b) Estimates for the static mixture model (left) and the dynamic Markov switching model (right)

Component	Static mixture model		Dynamic Markov switching model					
	Precision	Weight	Precision	Initial weight	Transition probability from [row] to ...			
					Level 0	Level 1	Level 2	QRE
Level 0	0	0.047	0	0.029	0.99	0	0	0
	(–)	(–)	(–)	(–)	(–)	(0)	(0)	(0)
Level 1	0.518	0	0.046	0.492	0	0.241	0.759	0
	(1.391)	(0.001)	(0.0052)	(0.100)	(–)	(0.089)	(0.092)	(0.030)
Level 2	0.043	0.157	0.165	0.411	0	0.130	0.576	0.294
	(0.018)	(0.045)	(0.0639)	(0.105)	(–)	(0.042)	(0.063)	(0.056)
QRE	0.091	0.795	1.315	0.068	0	0.179	0.079	0.741
	(0.008)	(0.050)	(0.227)	(0.062)	(–)	(0.060)	(0.043)	(0.068)
α	0.146		0.0221					
	(0)		(0.0135)					
BIC	692.23		646.74					
(LL)	(–677.26)		(–606.11)					

Note: The estimated models are mixtures of levels 0–2 and QRE. For each component, the table lists the estimated precision parameter (λ), the initial weight ρ_c , and as for the Markov switching model, also the transition probabilities $\omega_{c,c'}$ for transitions from component c to c' . The probabilities are derived from logistic transition propensities $\psi_{c,c'}$, through $\omega_{c,c'} = \exp\{\psi_{c,c'}\} / \sum_{c'} \exp\{\psi_{c,c'}\}$, and these propensities are the actual objects of estimation (to avoid boundary solutions; also, $\psi_{c,0}$ were normalized to zero). α is the social preference parameter. The standard errors given in parentheses are bootstrapped.

(c) The proportions of types implied by the Markov switching model and the reinforcement model, throughout the experiment and ad infinitum

	Segment								
	1	2	3	4	5	6	7	8	∞
<i>Markov switching rule learning</i>									
Level 0	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
Level 1	0.492	0.184	0.155	0.159	0.162	0.164	0.165	0.166	0.166
Level 2	0.411	0.616	0.508	0.435	0.401	0.387	0.381	0.379	0.379
QRE	0.068	0.172	0.308	0.378	0.408	0.420	0.425	0.427	0.428
<i>Reinforcement rule learning</i>									
Level 0	0.187	0.109	0.107	0.019	0.03	0.001	0.002	0	
Level 1	0.296	0.345	0.346	0.375	0.378	0.34	0.36	0.275	
Level 2	0.195	0.148	0.147	0.108	0.113	0.1	0.101	0.1	
QRE	0.321	0.399	0.4	0.498	0.479	0.559	0.537	0.624	

Markov switching: The proportions of types, i.e. component weights, in round 1 are the estimated initial weights (Table 9b). The weights for the subsequent rounds are computed using the transition matrix. The extrapolation toward infinity has been computed assuming the weight of level 0 remains at 0.028, as estimated for all rounds in the experiment. Proportions do not always add up to 1 due to rounding.

Reinforcement: These are the average probabilities (averaged across subjects) to use the four alternative rules in the various segments of the experiment. Extrapolation toward infinity is infeasible here.

they are less likely to maintain their prior that all subjects stick to level 2 or to equilibrium behavior respectively, and as a result may change their beliefs about the other subjects' reasoning modes and revert to level 1.

Result 12 (Belief rule dynamics). *The Markov switching rule learning model that approaches equilibrium along the level-k path best explains the data, outperforming the static mixture model with constant rules, the Markov switching model along the QRE path, and the reinforcement-based rule learning model. There is a systematic transition from level 1 to level 2 to QRE, followed by reversions from level 2 or QRE to level 1.*

Reinforcement learning is known to apply well to action-based learning. Our result that rule learning is better modeled using Markov switching than using reinforcement learning suggests that rule-based learning exhibits higher inertia, more specific state transitions, and less experimentation than action-based learning. This matches with the observation that the

Markov model defined over eight segments fits better than that over four segments, while the opposite is true for the reinforcement model as rule reinforcement does not fit here.¹⁹

These models do not simply fit better, but also yield qualitatively better results. Table 9c illustrates this. It provides the proportions of behavioral types across the segments of the experiments. As discussed, very few subjects (6.8%) start out as QRE subjects. Instead, level 1 has the highest initial weight (49.2%), but those subjects quickly adopt level 2 behavior, and this in turn leads to an increase of QRE behavior in subsequent segments. In the second half of the experiment, through segments 5–8, the proportions of behavioral types are stable, and by the end, they are virtually converged to their fixed point, the unit eigenvector of the transition matrix. We can therefore conclude that, in support of Hypothesis 8, subjects have converged to a *steady state*, an inter-temporal state in which type distributions are constant ad infinitum. In the steady state, there is a type distribution, with 2.9%, 16.6%, 37.9% and 42.6% of level 0, 1, 2 and QRE types respectively; and there are type switches from level 1 to level 2, from level 2 to QRE, and from level 2 and QRE to level 1, maintaining the proportions for each type ad infinitum. When subjects observe opponents playing as higher (lower) types than previously assumed, they switch to higher (lower) types. A distribution of different types implies corresponding switches to different types. The type distribution is stable because the probabilities of switching to and from a type are equal in the steady state.

Result 13 (Convergence). *The steady state contains a mixture of levels 0, 1, 2, and QRE types (2.9%, 16.6%, 37.9%, and 42.6%, respectively), sustained by continuing switches from levels 1 to 2, from level 2 to QRE, and reversions to level 1.*

In contrast, note the corresponding type proportions across segments implied by the estimated reinforcement model; see also Table 9c. As discussed above, the main structural difference is that by the reinforcement model, the belief rule employed in the subsequent segment is assumed to be independent of the belief rule employed in the current segment. That is, a subject playing level 0 or level 1 currently is assumed to be as likely to play QRE in the next segment as a subject playing QRE currently. A comparison of the BIC's in Table 9a shows that this “independence assumption” is rejected significantly in favor of the Markov model. An implication of assuming such independence nonetheless is that entirely different type proportions are estimated. The residual component (level 0) would appear to be rather large initially (18.7%) but disappears eventually, level 2's weight seems rather small (and also decreasing) across the eight segments, and the population would appear to consist predominantly of level 1 and QRE types. As the difference to both initial and eventual type proportions implied by the Markov model is large, this illustrates that making the invalid independence assumption distorts results substantially.

7. Discussion and conclusion

This paper analyzed strategic choice in an experiment where subjects played dynamic “club games” with incomplete information and varying parameters. The club game is an experimental tool to separate boundedly rational types and to analyze learning as behavior converges to a steady state. For the limit points of level- k and QRE are unique, as opposed to those of other refinement concepts, and the club game enables us to study strategic decision making in a simple, intuitive environment where beliefs in the rationality and precision of others are of primary relevance. We proposed a Markov switching rule learning model to understand individual transitions between belief rules. This allowed us to investigate the relevance of quantal response and level- k reasoning in the learning dynamics.

Our analysis of level- k thinking in dynamic games contributes to a literature that has been focused on the initial play in normal form games (e.g. Costa-Gomes et al., 2001; Crawford and Iriberri, 2007). A few exceptions have applied level- k to two-stage games (Johnson et al., 2002; Stahl and Haruvy, 2008) and signaling games (Kawagoe and Takizawa, 2009). Kawagoe and Takizawa (2009) found that level- k explains their signaling game data better than equilibrium concepts including QRE. In their games, competing refinement concepts are theoretically plausible, and the limit points of QRE and level- k are different either from those of other refinement concepts or from each other. In club games, in contrast, level- k and QRE are analyzed without interference of equilibrium selection or other refinement concepts. Similarly, a small but growing literature applies these concepts to understand learning after initial play, e.g. level- k with increasing levels (Stahl, 1996) and QREs with increasing precision (see also Turocy, 2005). Most recently, Haruvy and Stahl (2012) showed that rule learning between games exhibits patterns different from those within games. Specifically, they found that subjects switched from non-belief based strategies (“herding”) to belief based ones (“level- n ”). In our games, herding is less relevant due to the heterogeneity of players, which allows us to focus on transitions between level- k and QRE.

Our basic results underscored the qualitative compatibility of the data with level- k and QRE. The sensitivity of behavior to incentives and mover positions are largely consistent with QRE and iterated logit response. The observed sensitivity of behavior to the required level of reasoning indicates that an explanation that neglects level- k reasoning would be incomplete. These findings correspond with those of many previous studies, such as those cited above, and are further underlined by the result that static mixture models including logit equilibrium and level- k components outperformed singular equilibrium models EFR and QRE. Throughout, we controlled for social preferences, captured best in our context with a one-parametric model of conditional spite, jointly with quantal response and level- k reasoning in order to avoid mis-specification of the

¹⁹ The less segments, the better in this case.

learning process. The relevance of conditional spite turned out to be small, which may depend on the experimental parameters we used, and none of our key results depend on its inclusion in the model.²⁰ Our key findings can be summarized as follows:

1. Individual beliefs on the level of reasoning of other players do not stay constant with experience: our Markov switching rule learning model allowing for level- k transitions outperforms static mixture models, which assume constant belief rules. Subjects learn and change their beliefs of other players' over time.
2. The Markov switching rule learning model allowing for level- k transitions outperforms other models including the reinforcement rule learning model and a Markov switching model allowing only for changes in precision. Thus, subjects advance to equilibrium via systematic transitions along the level- k learning path.
3. Belief rule changes at the individual level involve systematic transitions from level 1 to level 2 to QRE, and reversions of level 2 and QRE types to level 1, arguably after observing opponents deviating from their expectations. These transitional dynamics are stationary, leading to a steady state with constant type distributions that is sustained with continuing type transitions at the individual level.
4. The steady state of the *type distribution* is the unit eigenvector of the transition matrix. In this steady state, the population consists of 2.9%, 16.6%, 37.9% and 42.6% of level 0, 1, 2 and QRE types respectively. Throughout the second half of our experiment, our subject pool is very close to the steady state.

The Markov switching rule learning model fits well also in absolute terms, and much better than the static mixture model that assumed a time-constant mode of reasoning. Modes of reasoning were not individually constant and individual transitions should thus generally be controlled for in such analyses. The Markov model also fits significantly better than the reinforcement model. The analysis does not pick up on subjects strategically experimenting with rules; subjects understood and applied rules systematically. Transitions were largely systematic and monotonic, transitioning from level 1, through level 2, to QRE, but movements from QRE and level 2 back to level 1 did occur.

A plausible interpretation is that after subjects observed a deviating opponent, maintaining equilibrium beliefs was then objectively implausible, and thus this cyclical pattern would result. This finding relates to that of [Johnson et al. \(2002\)](#), who applied level- k to ultimatum bargaining, and found that while backward induction is quickly learned, those capable of equilibrium reasoning do not necessarily apply it when faced with less experienced co-players.

The equilibrium state we find contained all three major modes of reasoning (levels 1, 2, and QRE), and individual beliefs were not constant. Our dynamic model uncovered a dimension neglected by the static model, namely the transition of player beliefs that occur over time. Compared to the static model, the dynamic model obtained a better fit and hence more reliable estimates of component weights for the respective behavioral types. The static model is thus mis-specified, and in our analysis, static and dynamic models arrive at substantively different qualitative results. The same is potentially inherent in cognitive hierarchy studies that do not model transitions in the cognitive hierarchy. Further research may therefore use Markov switching models to examine how transition rules and steady state depend on the game in question, and which events trigger rule switching.

Acknowledgments

We thank three anonymous reviewers, an advisory editor, and Thomas Palfrey for very helpful comments. We are grateful to Friedel Bolle, Avinash Dixit, Alexia Gaudeul, Morten Hviid, David Ong, and participants to workshop presentations at the Universities of East Anglia, Keele, Nottingham, and York, and conference presentations in Rome (Economic Science Association Meeting) and Tokyo (Econometric Society Meeting), for advice and encouragement. We also thank Jan Heufer, Marta Sernec, Julia Lähnert, and Claudia Vogel for the experimental assistance. The usual disclaimer applies. The financial support of the Deutsche Forschungsgemeinschaft (Grants BO 747/8 and BR 4648/1), the ESRC (NIBS Grant ES/K002201/1), and Nottingham University Business School are gratefully acknowledged. The authors can be contacted for access to the data.

Appendix A. Relegated definitions and proofs

The set of possible move sequences is denoted as $\mathcal{R}$, and any particular move sequence $R \in \mathcal{R}$ is a one-to-one function $R: N \rightarrow \{1, \dots, n\}$ where $R(i)$ denotes the position of $i \in N$. The move sequence is chosen randomly by Nature according to a probability distribution $P \in \Delta(\mathcal{R})$ which is common knowledge and assigns a positive probability to every $R \in \mathcal{R}$.

Histories of actions are duples $(R, h) \in \mathcal{R} \times H$ for $H := \bigcup_{t=0}^n \{0, 1\}^t$, where $R \in \mathcal{R}$ is the move of nature and $h \in H$ is the possibly empty list of the decisions of the players who moved already. Thus, h_t is the action of $R^{-1}(t)$. Typical information sets are denoted as $I \in \bar{I}$, and the subset of $\bar{I}$ where player i is to move is $\bar{I}_i$. The strategy of player i is a function $\sigma_i: \bar{I}_i \rightarrow [0, 1]$, and $\sigma_i(t, k)$ denotes the probability that i joins in the information set (i, t, k) . Players cannot distinguish

²⁰ The confluence of iterative logit response and social preferences has been largely unexplored in the literature. One exception is [Gneezy \(2005\)](#), who added altruism to the cognitive hierarchy model of [Camerer et al. \(2004\)](#) to better explain behavior in first price auctions.

histories of actions that lead to the same information set. The belief of player i that the history (R, h) applies when he finds himself in the information set (i, t, k) is denoted as $\mu_i(R, h|t, k)$. Beliefs are updated according to Bayes' Rule, and in our case they can be expressed as follows. Fix a strategy profile σ . If the probability that $j \in N$ chooses action $a_j \in \{0, 1\}$ in (j, t, k) is denoted as $\sigma_j(t, k|a_j)$ for the moment, the probability that history (R, h) results is (a priori)

$$\Pr(R, h) = P(R) * \prod_{t' < t} \sigma_{R^{-1}(t')} \left(t', \sum_{t'' < t'} h_{t''} \right) (h_{t'}). \quad (\text{A.1})$$

The belief system of $i \in N$ satisfies Bayes' Rule if for all $R \in \mathcal{R}$, $t \in \{1, \dots, n\}$, $k < t$, and all $h \in \{0, 1\}^{t-1}$ such that $\sum_{t' < t} h_{t'} = k$,

$$\mu_i(R, h|t, k) * \sum_{\tilde{R} \in \mathcal{R}} \Pr(\tilde{R}, h) = \Pr(R, h). \quad (\text{A.2})$$

Given σ , let $\pi_i(\sigma)$ denote the expected payoff of $i \in N$. In addition, let $\pi_i(a_i, \sigma|R, h)$ denote the expected payoff of i if he chooses action $a_i \in \{0, 1\}$ after history (R, h) , where everything else is played according to σ . Sequential rationality is satisfied if, for all information sets (i, t, k) ,

$$\sigma_i(t, k) \in \arg \max_{s \in [0, 1]} \sum_{(R, h)} \mu_i(R, h|t, k) * [s * \pi_i(1, \sigma|R, h) + (1 - s) * \pi_i(0, \sigma|R, h)]. \quad (\text{A.3})$$

Using this notation, the unique limiting logit equilibrium (Proposition 2) can be stated as

$$\sigma_i(t, k)(1) = \begin{cases} 1, & \text{if } k \geq i - (n - t) - 1 \\ 0, & \text{otherwise.} \end{cases} \quad (\text{A.4})$$

Proof of Proposition 1. Assume an equilibrium exists where some $i \in N$ does not join with probability 1 in an information set that is reached with positive probability along the path of play. Then there also exists an equilibrium where some $i \in N$ does not join with probability 1 in all information sets along the path but all $j < i$ do so (note that $i = 1$ may apply, in which case the set of $j < i$ is empty). Fix any information set along the equilibrium path where i does not join with probability 1. By assumption, all $j < i$ join with probability 1 if i does not join in this information set. This implies, by Eqs. (1) and (4), that i is best off not to join only if he believes that less than $i - 1$ opponents join overall if he joins in that information set, which can be satisfied only if some player $j < i$ does not join if i joins in that information set. Hence $i > 1$, i joins with probability 0 in this information set (according the assumed equilibrium), and at least one $j < i$ has not moved yet (otherwise, $i - 1$ opponents would have joined already).

First, consider the case that i 's information set is reached through a history of actions where exactly one $j < i$ has not moved yet. In this case, if i joins, then j will have to move in an information set where at least $i - 1$ opponents have already joined, which implies that j joins by sequential rationality (contradicting the assumption that he does not join with probability 1). Second, consider the case that i 's information set is reached through a history of actions where more than one $j < i$ has not moved yet. In this case, by the assumption that all move sequences have positive probability, i 's joining in this information set does not imply that all of the succeeding opponents find themselves in information sets off the equilibrium path. Regardless of the actual move sequence, only the lastly moving player of the $j < i$ can find himself in an information set off the path. All other $j < i$ join because they are in information sets compatible with the equilibrium path, and the lastly moving $j < i$ joins by sequential rationality. Hence, all $j < i$ join even if i joins, which yields the contradiction. $\square$

Definition 14. A strategy profile σ is a *logit equilibrium* for $\lambda \geq 0$ if for all (i, t, k) , $\sigma_i^r(t, k) = \exp\{\lambda^r * \pi_i(1, \sigma|t, k)\} / (\exp\{\lambda^r * \pi_i(0, \sigma|t, k)\} + \exp\{\lambda^r * \pi_i(1, \sigma|t, k)\})$, where $\pi_i(a_i, \sigma|t, k)$ is i 's expected payoff under σ if he deviates to $a_i \in \{0, 1\}$ with probability 1 in (i, t, k) conditional on reaching the information set (i, t, k) . The profile σ is a *limiting logit equilibrium* if there exists a sequence (λ^r) converging to ∞ and a sequence (σ^r) converging to σ such that for all r , σ^r is a logit equilibrium for λ^r .

Proof of Proposition 2. By Theorem 2 of McKelvey and Palfrey (1998), all limiting logit equilibria (LLEs) correspond with strategies of sequential equilibria as λ approaches infinity. By Proposition 1 this implies that all players join in all information sets along the path of play, and by dominance arguments, it implies that they join with probability 1 in information sets (i, t, k) where $k \geq i - 1$ and with probability 0 in information sets (i, t, k) where $k + n - t < i - 1$. The following shows that they join with probability 1 in the remaining information sets. Let σ be an LLE as defined above (i.e. the limit of QREs as λ approaches infinity). First, we show that all players $i < n$ join with probability 1 in all information sets (i, t, k) where $k = t - 2$ (i.e. where exactly one opponent had chosen not to join). All players join with probability 1 along the path of play, i.e. in all information sets (i, t, k) where $k = t - 1$. Hence, as λ tends to infinity, the expected payoff of i in such an information set from joining equates with $p_i(1, n - 1)$, and the expected payoff from not joining equates with $p_i(0, n - 1)$. Hence, in the limit, the probability that i does not join when $k = t - 1$ is

$$1 - \sigma_i(t, k) = \frac{\exp\{\lambda \cdot p_i(0, k)\}}{\exp\{\lambda \cdot p_i(0, k)\} + \exp\{\lambda \cdot p_i(1, k)\}} = \frac{\exp\{\lambda \cdot 0\}}{\exp\{\lambda \cdot 0\} + \exp\{\lambda \cdot [p_i(1, k) - p_i(0, k)]\}}$$

and thus for all $i < j \leq n$, the following is a consequence of Eq. (3) as λ tends to infinity.

$$\frac{1 - \sigma_i(t, k)}{1 - \sigma_j(t, k)} = \frac{\exp\{\lambda \cdot 0\} + \exp\{\lambda \cdot [p_j(1, k) - p_j(0, k)]\}}{\exp\{\lambda \cdot 0\} + \exp\{\lambda \cdot [p_i(1, k) - p_i(0, k)]\}} \xrightarrow{\lambda \rightarrow \infty} 0 \quad (\text{A.5})$$

Speaking loosely, player i is infinitely less likely (as $\lambda \rightarrow \infty$) to deviate from the equilibrium path than j if $i < j$. Hence, in all information sets (i, t, k) with $k = t - 2$ and $i < n - 1$, i believes that with probability 1 it was player $j = n$ who deviated from the equilibrium path, and hence i is best off joining, and uniquely so by Eq. (2).

In a similar manner, we can show that all $i < n - 1$ are uniquely best off joining in all information sets (i, t, k) where $k = t - 3$, i.e. when two opponents have already chosen not to join. All players $i < n - 1$ believe that with probability 1 this must have been the players $n - 1$ and n , and hence all $i < n - 1$ are uniquely best off joining. By logical induction, the proposition can thus be proved for all information sets. $\square$

Definition 15 (*Extensive form rationalizability*). Let $\bar{I}_i(\sigma_{-i})$ denote the set of $I \in \bar{I}_i$ that are reached with positive probability under a given σ (which is independent of his own strategy). Similarly, for any $\Sigma'_{-i} \subseteq \Sigma_{-i}$, let $\bar{I}_i(\Sigma'_{-i})$ denote the $I \in \bar{I}_i$ such that there exists $\sigma_{-i} \in \Sigma'_{-i}$ for which $I \in \bar{I}_i(\sigma_{-i})$. The payoff of $i \in N$ conditional on information set $I \in \bar{I}_i$ being reached is denoted $\pi_i(\sigma|I)$. Now define $\Sigma_i^1 = \Sigma_i$ for all $i \in N$, and for all t ,

$$\Sigma_i^{t+1} = \left\{ \sigma_i \in \Sigma_i^t \mid \forall I \in \bar{I}_i(\Sigma_{-i}^t) \exists \sigma_{-i} \in \Sigma_{-i}^t: I \in \bar{I}_i(\sigma_{-i}) \text{ and } \sigma_i \in \arg \max_{\sigma'_i \in \Sigma_i^t} \pi(\sigma'_i, \sigma_{-i} | I) \right\}. \quad (\text{A.6})$$

The strategies in $R_i = \bigcap_{t \geq 1} \Sigma_i^t$ are extensive form rationalizable for player i . Note that this definition has been simplified suitably to fit our context. In relation to Pearce's Definition 9, point (vi) simplifies to a conditional payoff calculation. Besides, as each player moves exactly once along any path of play, points (ii) and (iii) become redundant (which allows us to skip the explicit notation of conjectures).

Proof of Proposition 3. For any information set (i, t, k) , let $\bar{n}(t, k) = k + (n - t) - 1$ denote the number of opponents that will have joined eventually if all later players join with probability 1. After the first round of eliminating strategies, player i joins with probability 1 in all information sets (i, t, k) where $k \geq k_i^*$, and he joins with probability 0 if $\bar{n}(t, k) < k_i^*$. As for player 1, this is equivalent to the claimed strategy Eq. (A.4). As for players $i > 1$, this includes all information sets where i is claimed not to join. We have to show that all $i > 1$ join with probability 1 in the remaining information sets as the number of iterations τ approaches infinity. Following Definition 15, let Σ_i^τ denote the set of i 's strategies that are not eliminated prior to round $\tau \geq 2$. Moreover, for all information sets (i, t, k) , let $A_{\tau,i}(t, k)$ denote i 's set of rationalizable actions in (i, t, k) and round τ . Assume that the following assumptions hold true in round τ of the induction (noting that they are satisfied in round $\tau = 2$, i.e. after round 1). (A1) i 's strategy set is the product set of the action sets $A_{\tau,i}$, i.e. $\Sigma_i^\tau = \prod_{I \in \bar{I}_i} A_{\tau,i}(I)$. (A2) For all players $i < \tau$ and all (t, k) , $A_{\tau,i}(t, k) = \{1\}$ if $\bar{n}(t, k) \geq k_i^*$ and $A_{\tau,i}(t, k) = \{0\}$ if $\bar{n}(t, k) < k_i^*$. (A3) For all $i \geq \tau$ and all (t, k) , $A_{\tau,i}(t, k) = \{1\}$ if $k \geq k_i^*$, $A_{\tau,i}(t, k) = \{0\}$ if $\bar{n}(t, k) < k_i^*$, and either $A_{\tau,i}(t, k) = \{1\}$ or $A_{\tau,i}(t, k) = [0, 1]$ if $\bar{n}(t, k) \geq k_i^* > k$. We claim that if the induction assumptions (A1)–(A3) are satisfied in round τ , then they are also satisfied in round $\tau + 1$. It follows that (A1)–(A3) hold true for all $\tau \geq 2$, hence that (A2) holds true for $\tau \geq n$, which proves the proposition.

Let $\bar{I}'_i \subseteq \bar{I}_i$ denote the subset of i 's information sets where $A_{\tau,i}(I) = [0, 1]$. By (A3), $\bar{n}(r, h) \geq k_i^*$ holds true in all $(i, t, k) \in \bar{I}'_i$. Fix $(i, t, k) \in \bar{I}'_i$, and define an ordering $R \in \mathcal{R}$ such that $R(i) = t$ and $R^{-1}(t') > i$ for as many $t' < t$ as possible, given (t, k) , and where these players $j > i$ move initially. As a consequence of $(i, r, h) \in \bar{I}'_i$ and $A_{\tau,i}(i, t, k) = [0, 1]$, it can be shown that for all players $j > i$ that move prior to i under the ordering R , $A_{\tau,j}(t', k') \ni \{0\}$ applies in their respective information sets. By (A1), this implies that there exists $\sigma_{-i} \in \Sigma_{-i}^t$ such that $\sigma_j(t', k') = 0$ for all these $j > i$ and the respective information sets. Hence, there exists some σ_{-i} such that information set $(i, t, k) \in \bar{I}'_i$ will be reached if ordering R is chosen. The ordering R is chosen with positive probability, and thus, any $(i, t, k) \in \bar{I}'_i$ can still be reached along the path of play (given Σ^τ).

An implication of (A2) is that all players $j < \tau$ join with probability 1 along all paths of play that are compatible with Σ^τ . Hence, for any $\sigma \in \Sigma^t$ and any move order R , at least $\tau - 1$ players join. By (A1) this applies regardless of how player $i = \tau$ moves in any information set $(i, t, k) \in \bar{I}'_i$. For player $i = \tau$, this implies by Eqs. (1) and (4) that player i is uniquely best off joining in any $(i, t, k) \in \bar{I}'_i$. Since all of those information sets can be reached along the path of play under Σ^τ in round τ , this confirms (A2) for $\tau + 1$. (A3) follows, because for all $j > \tau$ in information sets (j, t, k) where $\bar{n}(t, k) \geq k_j^* > k$, the equilibrium prediction $A_{\tau,j}(t, k) = \{1\}$ is obviously robust to elimination, and if any other strategy is eliminated in such an information set, then all but the equilibrium prediction get eliminated (since the action sets are binomial in every information set). Finally, (A1) applies also for $\tau + 1$ since every player moves exactly once for every move order. $\square$

Appendix B. Estimation of Markov switching models

The following describes the forward–backward algorithm to compute the likelihood and the EM algorithm to estimate the Markov switching model. For further information, see e.g. [Bilmes \(1998\)](#).

Notation. Components $c \in C$ with initial weights ρ_c , periods $t = 1, \dots, T$, state at t is an unobserved random variable C_t , transition probabilities $\omega_{cc'} = \Pr(C_t = c' | C_{t-1} = c)$, parameters θ_{ct} for component c in period t (parameters may be time dependent). State-paths are $\kappa \in C^T$ and have probability $\Pr(\kappa)$, conditional on ρ, ω . Subject s with observations o_{st} in period t . Probability of o_{st} conditional on inclusion in $c \in C$ in period t is denoted as $\Pr(o_{st} | \theta, c, t)$. The log-likelihood is thus

$$LL(\rho, \omega, \theta | O) = \sum_{s \in S} \ln \sum_{\kappa \in C^T} \Pr(\kappa) \cdot \prod_{t \leq T} \Pr(o_{st} | \theta, \kappa(t), t).$$

Expectation. Given (ρ, ω, θ) , compute the posterior probability of component inclusion for all subjects s and all periods t , using the forward–backward algorithm.

The probability of observing the partial sequence $o_{s1}, \dots, o_{st}$ ending up in state c at t ,

$$\alpha_{sc}(t) = \Pr(O_{s1} = o_{s1}, \dots, O_{st} = o_{st}, C_t = c),$$

is efficiently defined as

$$\alpha_{sc}(1) = \rho_c \Pr(o_{s1} | \theta, c, 1) \quad \alpha_{sc'}(t+1) = \left[\sum_{c \in C} \alpha_{sc}(t) \omega_{cc'} \right] \Pr(o_{s,t+1} | \theta, c', t+1).$$

The probability of the ending partial sequence $o_{t+1}, \dots, o_T$ conditional on starting in c at t ,

$$\beta_{sc}(t) = \Pr(O_{s,t+1} = o_{s,t+1}, \dots, O_{sT} = o_{sT}, C_t = c),$$

is efficiently defined as

$$\beta_{sc}(T) = 1 \quad \beta_{sc}(t) = \sum_{c' \in C} \omega_{cc'} \Pr(o_{s,t+1} | \theta, c', t+1) \beta_{sc'}(t+1).$$

The probability of observing the sequence $(o_{s1}, \dots, o_{sT})$ unconditional of states is

$$\Pr(o_s) = \sum_{c \in C} \alpha_{sc}(T) = \sum_{c \in C} \beta_{sc}(1) \rho_c \Pr(o_{s1} | \theta, c, 1).$$

The probability of being in state c at time t given o_s , $\gamma_{sc}(t) = \Pr(C_t = c | o_s)$, is

$$\gamma_{sc}(t) = \frac{\Pr(O, C_t = c)}{\Pr(O)} = \frac{\Pr(O, C_t = c)}{\sum_{c' \in C} \Pr(O, C_t = c')} = \frac{\alpha_{sc}(t) \beta_{sc}(t)}{\sum_{c' \in C} \alpha_{sc'}(t) \beta_{sc'}(t)}.$$

The probability of s being in state c at t and in c' at $t+1$, $\zeta_{sc'}(t) = \Pr(C_t = c, C_{t+1} = c' | o_s)$, can be derived similarly.

$$\begin{aligned} \zeta_{sc'}(t) &= \Pr(C_t = c, C_{t+1} = c' | o_s) = \frac{\Pr(C_t = c | o_s) \Pr(o_{t+1}, \dots, T, C_{t+1} = c' | C_t = c)}{\Pr(o_{t+1}, \dots, T | C_t = c)} \\ &= \frac{\gamma_{sc}(t) \omega_{cc'} \Pr(o_{s,t+1} | \theta, c') \beta_{sc'}(t+1)}{\beta_{sc}(t)} \end{aligned}$$

Maximization. For all $c \in C$ and $t \leq T$, define

$$LL_{ct}(\theta') = \sum_{s \in S} \gamma_{sc}(t) \cdot \ln \Pr(o_{s,t} | \theta', c)$$

and optimize it to update the parameter estimates $\theta_{ct}^{+1} \in \arg \max_{\theta'} LL_{ct}(\theta')$. In case θ_{ct} affects the likelihoods of several components (such as precision of level 1 or an altruism weight for all levels), let $C' \subseteq C$ denote the affected components, optimize $\theta_{ct}^{+1} \in \arg \max_{\theta'} \sum_{c' \in C} LL_{c't}(\theta')$. Finally, update

$$\rho_c^{+1} = 1/n \cdot \sum_{s \in S} \gamma_{sc}(1) \quad \omega_{cc'}^{+1} = \frac{\sum_{s \in S} \sum_{t < T} \zeta_{sc'}(t)}{\sum_{s \in S} \sum_{t < T} \gamma_{sc}(t)}$$

and reiterate until convergence is achieved.

Appendix C. Experimental instructions and questionnaire

[This is the English version of the experimental instructions and questionnaire. The experimental instructions and questionnaire used in the actual experiment were in German, and are available from the authors upon request.]

General instructions

You are about to participate in an experiment on decision-making. The experiment is divided into five stages, and each stage is divided into rounds. During the experiment you will earn *experimental points*.

At the start you are assigned an *initial endowment* of 60 experimental points. Each experimental point you earn in the experiment is worth 10 cents.

At the end of the experiment, three winning rounds will be randomly chosen by the computer, two from stages 1 through 4, and one from stage 5. What you earn in the winning rounds will be added or subtracted from the initial endowment to determine your final winnings. You will not know which rounds are the winning rounds until the end of the experiment.

There are twelve participants in the experiment. There are four *types of participants*, which we label as P, Q, R and S. There are three participants of each type. You can find out which type you are by looking at the right-hand corner of your computer display. Participant types are only relevant for stages 1 through 4. Your participant type will stay the same throughout the experiment.

Your decision in stages 1 through 4

There are eight rounds in each of stages 1 through 4. Each round you are *matched* with three other participants (*coparticipants*) in the room, *one of each participant type*, as a *set of participants*. Therefore, each set of participants is always made of four people and always has a P participant, a Q participant, an R participant and an S participant. If for example you are an S participant, this means that your coparticipants will always be a P participant, a Q participant and an R participant; a similar reasoning applies if you belong to one of the other participant types.

Coparticipants are chosen randomly each round from each of the other participant types, and so it is very unlikely that you will be matched in the same set of participants as you move from one round to the next.

In each round *you, and your matched coparticipants, are each asked in turn to decide whether you would like to become a member of a club*. You have to choose between either “yes” or “no”. How much you earn from the round depends on three factors: (a) your action; (b) how many of your matched coparticipants choose “yes”; (c) your participant type.

Information is provided each round in the form of decision tables. An illustrative example of decision table (**not** used in the experiment) is as follows:

	Final number of other participants who choose “yes”			
	0	1	2	3
Yes	30	40	50	60
No	60	50	40	30

If this decision table were to apply to you, it would tell you the amount of experimental points that you would earn from choosing “yes” or “no”, depending on the number of other participants who choose “yes”. For example, if you chose “yes” and two other participants also chose “yes”, you would earn 50 experimental points; similarly, you would get 50 points if one other participant chose “yes” but you chose “no”.

Decision tables vary according to the participant type, and the computer screen is currently displaying those that are applicable to stage 1 for the P participant, Q participant, R participant and S participant. Everyone is being given exactly the same set of information as that provided in this sheet and on the computer screen.

Decisions are taken in turns. The order in which you and your coparticipants make decisions is determined at random each round. Click on Check to find out if it is your turn. A message will appear when it is your turn to make your decision. It will inform you of how many coparticipants have made their decisions and chosen “yes” so far in the round.

When it is your turn to decide, and you have made your decision, if you choose “yes”, click the “yes” button; if you choose “no”, click the “no” button. To confirm your decision, click on Confirm, and a message box will appear asking if you are sure of your decision. If so, click on OK in the message box, and then on Confirm again. If not, you may change your decision either by clicking on Cancel, and then entering your new decision. You may change your decision anytime before the 2nd click on Confirm.

After everyone has made their decision for the round, you will be told the final number of other participants who have chosen “yes” and your corresponding earning for the round. You may then move on to the next round by clicking on Continue.

The same set of decision tables is used for all the rounds in a stage; however, decision tables change across stages.

Before starting stage 1, we ask you to answer a brief questionnaire, with the only purpose of checking whether you have understood the instructions. Raise your hand when you have completed the questionnaire.

Many thanks for your participation to the experiment, and good luck!

Please raise your hand if you have any questions.

Questionnaire

1. “The set of participants is made by everyone in the room.” True or false? _____
2. “Since coparticipants are chosen randomly every round, there can be rounds in which my set of participants includes two Q participants and an S participant.” True or false? _____
3. “Each of my coparticipants is drawn from a different participant type, and a different one from the one I belong to.” True or false? _____
4. “The people who make my set of participants always remain the same round after round.” True or false? _____
5. “Each round coparticipants (in my set) need to decide in turn whether to join the same club.” True or false? _____
6. What participant type are you? _____

For the following questions consider the decision tables currently on the computer display.

7. How many experimental points would participant P earn if he or she chooses “yes” and none of his or her coparticipants does? _____
8. How many experimental points would participant S earn if he or she chooses “no” and one of his or her coparticipants chooses “yes”? _____
9. How many experimental points would participant R earn if he or she chooses “yes” and two of his or her coparticipants choose “no”? _____
10. How many experimental points would participant Q earn if he or she chooses “no” and all his or her coparticipants choose “yes”? _____

Please raise your hand when you have completed the questionnaire. Do not start making choices until an experimenter checks your answers.

Appendix D. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.geb.2014.03.002>.

References

- Anderson, S., Goeree, J., Holt, C., 1998. Rent seeking with bounded rationality: an analysis of the all-pay auction. *J. Polit. Economy* 106 (4), 828–853.
- Asheim, G., 2002. Proper rationalizability in lexicographic beliefs. *Int. J. Game Theory* 30 (4), 453–478.
- Battalio, R., Samuelson, L., Huyck, J., 2001. Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica* 69 (3), 749–764.
- Battigalli, P., 1997. On rationalizability in extensive games. *J. Econ. Theory* 74 (1), 40–61.
- Bernheim, B., 1984. Rationalizable strategic behavior. *Econometrica* 52 (4), 1007–1028.
- Bilmes, J.A., 1998. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Mimeo.
- Camerer, C., Hua Ho, T., 1999. Experience-weighted attraction learning in normal form games. *Econometrica* 67 (4), 827–874.
- Camerer, C., Ho, T., Chong, J., 2004. A cognitive hierarchy model of games. *Quart. J. Econ.* 119 (3), 861–898.
- Carlsson, H., van Damme, E., 1993. Global games and equilibrium selection. *Econometrica* 61 (5), 989–1018.
- Cooper, D., Van Huyck, J., 2003. Evidence on the equivalence of the strategic and extensive form representation of games. *J. Econ. Theory* 110 (2), 290–308.
- Costa-Gomes, M., Crawford, V., 2006. Cognition and behavior in two-person guessing games: an experimental study. *Amer. Econ. Rev.* 96 (5), 1737–1768.
- Costa-Gomes, M., Crawford, V., Broseta, B., 2001. Cognition and behavior in normal-form games: an experimental study. *Econometrica* 69 (5), 1193–1235.
- Crawford, V., Iriberri, N., 2007. Level- k auctions: can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica* 75 (6), 1721–1770.
- Dixit, A., 2003. Clubs with entrapment. *Amer. Econ. Rev.* 93 (5), 1824–1829.
- Erev, I., Roth, A., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Amer. Econ. Rev.* 88 (4), 848–881.
- Frankel, D., Morris, S., Pauzner, A., 2003. Equilibrium selection in global games with strategic complementarities. *J. Econ. Theory* 108 (1), 1–44.
- Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. Springer Verlag.
- Gaviria, A., Raphael, S., 2001. School-based peer effects and juvenile behavior. *Rev. Econ. Statist.* 83 (2), 257–268.
- Gneezy, U., 2005. Step-level reasoning and bidding in auctions. *Manage. Sci.* 51 (11), 1633–1642.
- Goeree, J., Holt, C., Palfrey, T., 2002. Quantal response equilibrium and overbidding in private-value auctions. *J. Econ. Theory* 104 (1), 247–272.
- Haruvy, E., Stahl, D., 2012. Between-game rule learning in dissimilar symmetric normal-form games. *Games Econ. Behav.* 74 (1), 208–221.
- Ho, T., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental “ p -beauty contests”. *Amer. Econ. Rev.* 88 (4), 947–969.
- Hyndman, K., Ozbay, E., Schotter, A., Ehrblatt, W., 2012. Convergence: an experimental study of teaching and learning in repeated games. *J. Europ. Econ. Assoc.* 10 (3), 573–604.
- Johnson, E., Camerer, C., Sen, S., Rymon, T., 2002. Detecting failures of backward induction: monitoring information search in sequential bargaining. *J. Econ. Theory* 104 (1), 16–47.
- Kajii, A., Morris, S., 1997. The robustness of equilibria to incomplete information. *Econometrica* 65 (6), 1283–1309.
- Kawagoe, T., Takizawa, H., 2009. Equilibrium refinement vs. level- k analysis: an experimental study of cheap-talk games with private information. *Games Econ. Behav.* 66 (1), 238–255.
- Kreps, D., Wilson, R., 1982. Sequential equilibria. *Econometrica* 50 (4), 863–894.
- Krolzig, H., 1997. *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer, Berlin.
- Kübler, D., Weizsäcker, G., 2004. Limited depth of reasoning and failure of cascade formation in the laboratory. *Rev. Econ. Stud.* 71 (2), 425–441.
- Kuran, T., 1998. Ethnic norms and their transformation through reputational cascades. *J. Legal Stud.* 27 (S2), 623–659.
- McFadden, D., 1976. Quantal choice analysis: a survey. *Ann. Econ. Soc. Meas.* 5 (4), 363–390.
- McFadden, D., 1984. Econometric analysis of qualitative response models. In: *Handbook of Econometrics*, vol. 2, pp. 1395–1457.

- McKelvey, R., Palfrey, T., 1995. Quantal response equilibria for normal form games. *Games Econ. Behav.* 10 (1), 6–38.
- McKelvey, R., Palfrey, T., 1998. Quantal response equilibria for extensive form games. *Exper. Econ.* 1 (1), 9–41.
- McKelvey, R.D., McLennan, A.M., Turocy, T.L., 2007. Gambit: software tools for game theory, version 0.2007.01.30. <http://econweb.tamu.edu/gambit>.
- McLachlan, G., Peel, D., 2000. *Finite Mixture Models*, vol. 299. Wiley-Interscience.
- Morris, S., Rob, R., Shin, H., 1995. p -dominance and belief potential. *Econometrica* 63 (1), 145–157.
- Myerson, R., 1978. Refinements of the Nash equilibrium concept. *Int. J. Game Theory* 7 (2), 73–80.
- Nagel, R., 1995. Unraveling in guessing games: an experimental study. *Amer. Econ. Rev.* 85 (5), 1313–1326.
- Pearce, D., 1984. Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52 (4), 1029–1050.
- Powell, M., 2008. Developments of NEWUOA for minimization without derivatives. *IMA J. Numer. Anal.* 28 (4), 649.
- Rogers, B., Palfrey, T., Camerer, C., 2009. Heterogeneous quantal response equilibrium and cognitive hierarchies. *J. Econ. Theory* 144 (4), 1440–1467.
- Schelling, T.C., 1960. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA.
- Schuhmacher, F., 1999. Proper rationalizability and backward induction. *Int. J. Game Theory* 28 (4), 599–615.
- Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory* 4 (1), 25–55.
- Selten, R., Stoecker, R., 1986. End behavior in sequence of finite prisoner's dilemma supergames: a learning theory approach. *J. Econ. Behav. Organ.* 7 (1), 47–70.
- Stahl, D., 1996. Boundedly rational rule learning in a guessing game. *Games Econ. Behav.* 16 (2), 303–330.
- Stahl, D., 1998. Is step- j thinking an arbitrary modelling restriction or a fact of human nature?. *J. Econ. Behav. Organ.* 37 (1), 33–51.
- Stahl, D., 2000. Rule learning in symmetric normal-form games: theory and evidence. *Games Econ. Behav.* 32 (1), 105–138.
- Stahl, D., Haruvy, E., 2008. Level- n bounded rationality in two-player two-stage games. *J. Econ. Behav. Organ.* 65 (1), 41–61.
- Stahl, D., Wilson, P., 1994. Experimental evidence on players' models of other players. *J. Econ. Behav. Organ.* 25 (3), 309–327.
- Stahl, D., Wilson, P., 1995. On players' models of other players: theory and experimental evidence. *Games Econ. Behav.* 10 (1), 218–254.
- Turocy, T., 2005. A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games Econ. Behav.* 51 (2), 243–263.
- Ui, T., 2001. Robust equilibria of potential games. *Econometrica* 69 (5), 1373–1380.
- Vuong, Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (2), 307–333.