Games and Economic Behavior 150 (2025) 71–92

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

# Taming selten's horse with impulse response[☆]

Tibor Neugebauer[a,*], Abdolkarim Sadrieh[b], Reinhard Selten[c,1]

[a] University of Luxembourg, Luxembourg
[b] University of Magdeburg, Germany
[c] University of Bonn, Germany

ARTICLE INFO

*JEL classification:*
C90
D53
D92
G02
G11
G12

*Keywords:*
Trembling-hand Perfect Equilibrium
Game of Selten's Horse
Learning Direction Theory
Impulse Response Dynamics
Quantal Response
Reinforcement Learning
Level-k

ABSTRACT

The paper experimentally examines the predictive power of the trembling-hand perfect equilibrium concept in the three-player Game of Selten's Horse. At first sight, our data show little support of the trembling-hand perfect equilibrium and rather favor the imperfect equilibrium. We introduce deterministic impulse response trajectories that converge on the trembling-hand perfect equilibrium. The impulse response trajectories are remarkably close – closer than the trajectories from a reinforcement learning model – to the observed dynamics of the game in the short run (50 periods). The quantal response approach also converges on the trembling-hand perfect equilibrium as the error rates decline, suggesting that the trembling-hand perfect equilibrium may be reached in the long run. In the long run (up to 250 periods), however, behavior seems to settle at a non-equilibrium distribution of strategies that rather supports efficient outcomes, instead of converging to the trembling-hand perfect equilibrium.

## 1. Introduction

Selten's *subgame perfection* (Selten 1965, 1975) has been a strikingly powerful refinement of the Nash equilibrium concept in the theory of extensive form games. However, to illustrate that not every intuitively unreasonable equilibrium point is excluded by the definition of subgame perfection, Selten (1975) proposed a numerical example which was later on referred to as *Selten's Horse* (Binmore 1987). Selten's Horse is a three-player game with no proper subgames. Every player has exactly one information set. Selten suggested the (*trembling-hand*) *perfect equilibrium* refinement (Selten 1973) along with a perturbation of the game to select a unique

[☆] Financial support by the University of Luxembourg is acknowledged (F2R-LSF-367-POL-09BCCM). The scientific research presented in this publication has also been given financial support by the National Research Fund of Luxembourg (11,585,355 & 16,991,801)

[*] Corresponding author.
*E-mail address:* tibor.neugebauer@uni.lu (T. Neugebauer).

[1] Reinhard Selten was involved in this research project until his untimely death, i.e., during the preparation of the research design, the hypotheses, the analysis of the short-term sessions, and the first full draft of the paper. We remain indebted to him for his incredible insight and creativity but take responsibility for any errors that remain in the paper despite our conscientious effort and the thorough examination by two excellent anonymous reviewers and the exceptional attention of Vince Crawford as editor. We also acknowledge the helpful comments from participants at seminars hosted by the University of Auckland, the University of Granada, the University of Innsbruck, the Royal Melbourne Institute of Technology, Monash University (Melbourne), Poznan University of Economics and Business, Luiss (Rome), and Tilburg University, as well as the feedback of attendees of the following conferences: Economic Design (Istanbul), ESA meeting (Tucson), GfeW meeting (Magdeburg).

https://doi.org/10.1016/j.geb.2024.11.014
Received 2 August 2024;
Available online 30 November 2024

equilibrium point. The perturbation of the game builds on the idea that each player makes mistakes with a small probability. The limiting equilibrium point on the perturbed game is a perfect equilibrium point. The perfect equilibrium concept, in general, and the perfect equilibrium point of the perturbed game, in particular, serve as a selection mechanism for situations with a multiplicity of equilibrium points.

In our study, we present experiments of the Game of Selten's Horse seeking empirical evidence in support of the trembling-hand perfect equilibrium. To our surprise, we observe very little support for the play of the trembling-hand perfect equilibrium strategies. Application of learning direction theory (Selten and Stoecker 1986; Selten and Buchta 1999) seems to capture the observed pattern of play much better when compared to the perfect equilibrium prediction. Curiously, the attraction point of learning direction dynamics, the impulse balance (Selten 2004), which we determine by examination of the impulse response dynamics (Ockenfels and Selten 2005; Selten and Chmura 2008; Goerg et al., 2016), is again contained in the set of perfect equilibrium points. The simulations of impulse response dynamics seem to closely reproduce the observed trajectories for most groups in the experiment with 50 periods ("short--run"). Based on this first experimental evidence, we tentatively conclude that the dynamics of play show a clear attraction to the perfect equilibrium, but that full convergence may take more than 50 periods.

However, in our second experiment, in which we extend the number of repetitions up to 250 periods ("long-run"), we do not find convergence to the perfect equilibrium as expected. Instead, we find that subjects learn to resist the attraction of the perfect equilibrium and tend to settle at strategy combinations that support high levels of mutual payoffs exceeding the trembling-hand perfect equilibrium. The prevalence of high levels of mutual payoffs are especially prevalent when subjects play the repeated game in fixed groups as "partners". In this long-run repeated game setting, the trajectories of the reinforcement learning model (Erev and Roth 1998) seem to provide a better fit than those of the impulse response. Similar to impulse response dynamics, the quantal response approach (McKelvey and Palfrey 1995) also suggests a convergence on the trembling-hand perfect equilibrium when error rates decline. The quantal response estimation indicates lower error rates in later than in earlier periods in the short-run experiment, thus, entertaining the hope that the trembling-hand perfect equilibrium may prevail with long repetitions. In the long-run experiment, however, behavior seems to settle at a non-equilibrium distribution of strategies that supports efficient outcomes or low level-k of strategic sophistication (Crawford, Costa-Gomes, Iriberri 2013) instead of converging to the trembling-hand perfect equilibrium.

The remainder of the paper is structured as follows. Section 2 introduces the Game of Selten's Horse and offers a discussion of the trembling-hand perfect and the imperfect equilibria. Section 3 describes our experimental design. In Section 4, we report our overall results and provide the static analyses of behavior and outcomes. In Section 5, we analyze the dynamics of play. We discuss learning direction theory, impulse response dynamics, and reinforcement learning trajectories, comparing the simulated trajectories to the data. Insights from alternative concepts, especially the quantal response model, the level-k model, and the joint payoff maximum, are also presented in this section. Section 6 provides an overview and a discussion of the findings and concludes the paper.

## 2. Theoretical considerations

*Selten's Horse* is depicted in Fig. 1. It is a three-player game with perfect recall, where every player has one information set. No proper subgames exist. Each player has two choices *L* and *R*. A strategy profile represents the actions of the players; e.g., (*R, L, R*) indicates that players 1 and 3 play *R* and player 2 plays *L*. Each pure strategy profile leads to a payoff triple; e.g., (*R, L, R*) leads to the payoff triple [4, 4, 0] where players 1 and 2 receive each a payoff of 4 and player 3 receives zero.

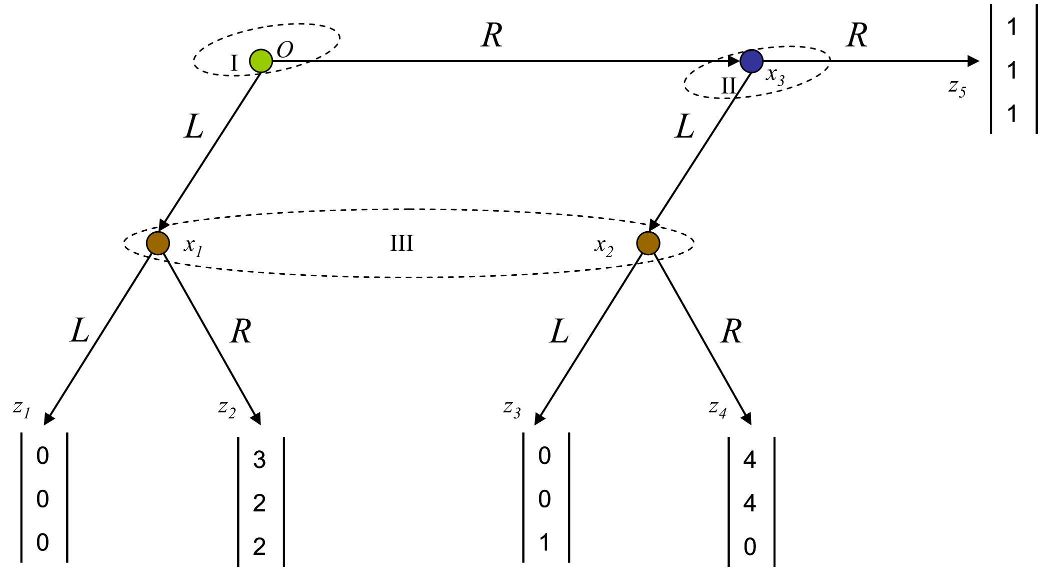


**Fig. 1.** Selten's horse.

Since each player has only two pure strategies, a *behavior strategy* of player $i$ can be characterized by the probability by which he or she selects $R$. Following Selten (1975), the symbol $p_i$ will be used for this probability. A combination of behavior strategies is represented by the strategy profile $(p_1, p_2, p_3)$.

The best response functions (1) in the Game of Selten's Horse are

$$p_1 = \begin{cases} 1, \\ [0,1] \\ 0 \end{cases} \quad \mathit{if}\ 4(1-p_2)p_3 + p_2 \quad \begin{matrix} > \\ = 3p_3 \\ < \end{matrix}$$

$$p_2 = \begin{cases} 1, \\ [0,1] \\ 0 \end{cases} \quad \mathit{if} \qquad p_3 \quad \begin{matrix} < \\ = 0.25 \\ > \end{matrix} \tag{1}$$

$$p_3 = \begin{cases} 1, \\ [0,1] \\ 0 \end{cases} \quad \mathit{if} \quad 2(1-p_1) \quad \begin{matrix} > \\ = p_1(1-p_2) \\ < \end{matrix}$$

There are two types of equilibria:

*Trembling-hand perfect equilibria* $p_1 = 1,\ p_2 = 1,\ 0 \le p_3 \le \frac{1}{4}$

*Imperfect equilibria* $p_1 = 0,\ \frac{1}{3} \le p_2 \le 1,\ p_3 = 1$

Selten (1975) proposed the trembling-hand perfect equilibrium refinement concept. The concept eliminates all imperfect equilibrium points in the Game of Selten's Horse by using a perturbed version of the game that selects a unique trembling-hand equilibrium point as its limit point. Let us first review the discussion of Selten (1975) of the imperfect equilibria, followed by the discussion of the trembling-hand perfect equilibria.

Imperfect equilibria are considered as unreasonable because of player 2′s choices. If players 1 and 3 play their imperfect equilibrium strategies, player 2′s expected payoff does not depend on the own strategy. Since player 2′s information set is not reached, any strategy – including any in the imperfect equilibrium strategy set – is a best response. In order to support the imperfect equilibrium strategies of the others, player 2 is required to choose a strategy from the set of imperfect equilibrium strategies, i.e., choose $R$ with a probability greater than one third. To see why it is unreasonable to expect that player 2 chooses to play $R$ if the own information set (node$x_3$ in Fig. 1) is reached, assume the following: The players believe a specific imperfect equilibrium point, e.g.,$(0, 1, 1)$, is the rational way to play Selten's Horse. When $x_3$ is reached this belief has been shown to be wrong. Player 2 has to take for granted that player 1 has chosen $R$. If player 2 believes that player 3 will choose $R$ according to the imperfect equilibrium point, then player 2′s best response is $L$ with a payoff of 4 instead of $R$ with a payoff of 1. The same reasoning also applies to the other imperfect equilibria. In contrast, in the trembling-hand perfect equilibria, player 3′s trembling-hand perfect equilibrium strategies maximize the own expected payoff, even if players 1 and 2 make mistakes so that player 3′s information set is reached, although it should not have been reached.

Selten (1975) formalizes the notion of players making small mistakes in his concept of the *perturbed game*. In the perturbed game, players play with "trembling hands," i.e., make mistakes with some very small probability$\varepsilon > 0$. Constructing a test sequence of $k$ perturbed games with $\varepsilon_k \to 0$ and $k \to \infty$, the trembling-hand perfect equilibrium is defined as the limit of the test sequence. Selten (1975) shows that in the Game of Selten's Horse, all trembling-hand perfect equilibria are limit points of test sequences of the perturbed game.[2]

## 3. Experimental design

To test the prediction of trembling-hand perfect equilibrium theory, we conducted six sets of experiments between 2010 and 2021, involving 3 × 2 between-subjects treatment conditions.[3] First, two sets of three computerized (Fischbacher 2007) experimental sessions with 50 periods of Selten's Horse were conducted employing a *strangers* treatment. Then, as a robustness check for our results, four additional sets of computerized experimental sessions with a maximum number of 250 periods of Selten's Horse followed: two sets of three sessions with strangers treatment and another two sets with *partners* treatment. All sessions were conducted using the same recruiting tools and procedures at the MaxLab of the University of Magdeburg. The first set of sessions ("short-run") were played for exactly 50 periods, each lasting about an hour. The second and third set of sessions ("long-run") were timed to take two hours, with the subjects knowing that the session would end at that time no matter how many periods were played by then.

The strangers treatment involved a random matching setup, where each session involved 27 subjects, split in 3 independent groups of 9 interacting subjects (3 subjects of each player type) who were randomly rematched in each period. Subjects maintained their player type throughout the entire session. Subjects were not informed on the details of the matching procedure, but they were informed that the likelihood of being rematched with the same two subjects in consecutive periods was small.

In the partners treatment we used a fixed matching setup, in which the 3 subjects knew that they would repeatedly interact with

[2] For further details and proofs see Selten (1975).

[3] As further explained in this section, subjects faced either an average-pay condition or a random-pay condition. They participated in one of three session types: short-run sessions with strangers matching, long-run sessions with strangers matching or long-run sessions with partners matching.

each other in all periods throughout the session. Hence, the partners treatment basically induces a finitely repeated game, presumably facilitating to coordinate on a specific outcome of the game.

To implement behavioral strategies and make them easily accessible to subjects, each period contained 100 automated plays of the Game of Selten's Horse for the three players. In every period, each subject of player type $i$ chose the relative frequency$f_i$ of playing $R$ in the 100 plays of that period, knowing that with the remaining frequency, $(1 - f_i)$, $L$ would be played. These choices represent our experimental implementation of the behavior strategies $p_i$ in the Game of Selten's Horse. The sequences of $R$ and $L$ actions in the 100 plays were randomly drawn (without replacement) for each player $i$ according to$f_i$. The outcomes of the 100 plays were determined by the action profiles resulting in the series of the three interacting players.

There are two standard implementations of mixed strategy payoffs in game theory experiments, which we apply as treatment variations; *random pay* and *average pay*. The players either receive (i) the average payoff of all plays or (ii) the payoff of one randomly selected play.[4] In our *Average Pay Condition,* the payment in a period is equal to the average payoff over the 100 plays of that period. In the *Random Pay Condition,* the period payment is equal to the outcome realized in one of the 100 plays of that period. Since each play counts equally under Average Pay, whereas only one play counts under Random Pay, the latter involves a much higher payoff variance than the former.

Subjects were provided with the same feedback in all treatment conditions. The outcomes in the 100 plays of a period were presented on the screen in a histogram that showed the observed frequencies of each outcome $z_1, \ldots, z_5$ of the game. The subjects were also shown the outcome of one particular play in each period. Subjects additionally received a record of their past period earnings and their total earnings.

## 4. General observations

For each matching protocol (short-run strangers treatment, long-run strangers treatment, and long-run partners treatment) we collected 18 independent observations, with 9 in Average Pay and 9 in Random Pay. In total, 162 subjects participated in the short-run sessions, earning an average payoff of € 16.10 with sessions completed within an hour, including instructions. In the long-run sessions, 162 subjects participated in the strangers treatment (9 independent matching groups with 9 subjects in each of the two payoff conditions) and 54 subjects in the partners treatment (9 groups of 3 subjects in each of the two payoff conditions). The long-run sessions were completed within two hours, including instructions. The long-run average payoffs for strangers were € 25.92, and for partners € 27.74. The number of periods that were played in the long-run sessions varied between 111 and 230 depending on the pace of the groups' interaction. Subjects received no show-up fee.

### 4.1. Overall distribution of outcomes

Fig. 2 shows the observed distribution of outcomes in the two payment conditions, corresponding to the notation $\{z_1, z_2, \ldots, z_5\}$ as in Fig. 1. We only find insignificant differences between outcome frequencies across the average-pay and random-pay conditions. Even the largest differences in the relative frequencies between Average Pay and Random Pay – as observed for the equilibrium outcomes $z_2$ $(0, \cdot, 1)$ and $z_5$ $(1, 1, \cdot)$ – are not significant.

Observed average earnings in the Average Pay and the Random Pay conditions are also rather similar. Compared to the trembling-hand perfect equilibria, we find that the average earnings of player 3 are close to this equilibrium prediction, but players 1 and 2 earn clearly in excess of this equilibrium prediction. Hence, even though the game structure provides incentives to select the trembling-hand perfect equilibrium, out-of-equilibrium play empirically seems to entail very few negative payoff effects. In contrast, all three players earn substantially less on average than in the imperfect equilibrium.

The average relative frequency of obtaining the imperfect equilibrium outcome $z_2$ is 0.734 in the partners treatment, whereas it is 0.409 in the long-run strangers and 0.368 in the short-run strangers treatment. In contrast, the trembling-hand perfect equilibrium outcome $z_5$ is only observed with an average frequency of 0.099 in the long-run partners treatment, of 0.156 in the long-run strangers treatment, and of 0.169 in the short-run strangers treatment. The differences in the relative frequencies of both imperfect and trembling-hand perfect equilibrium outcomes between partners treatment and strangers treatment are significant. Using the Mann-Whitney *U test*, we find p-values of 0.001 for the imperfect and of 0.002 for the trembling-hand perfect equilibrium outcomes when comparing the long-run partners treatment to the long-run strangers treatment and of 0.000 for the imperfect and of 0.004 for the perfect equilibrium outcomes when comparing the long-run partners treatment to the short-run strangers treatment.

***Observation 1a:*** The distributions of game outcomes show no significant treatment effects concerning the average-pay versus random-pay condition or short-run versus long-run treatments.

***Observation 1b:*** The distributions of game outcomes show significant treatment effects between partners and strangers. The imperfect equilibrium outcome is reached more frequently and the perfect equilibrium outcome less frequently in the partners treatment than in the strangers treatment.

[4] Friedman and Oprea 2012 use similar payoff protocols studying mixed strategies in a prisoner's dilemma game.

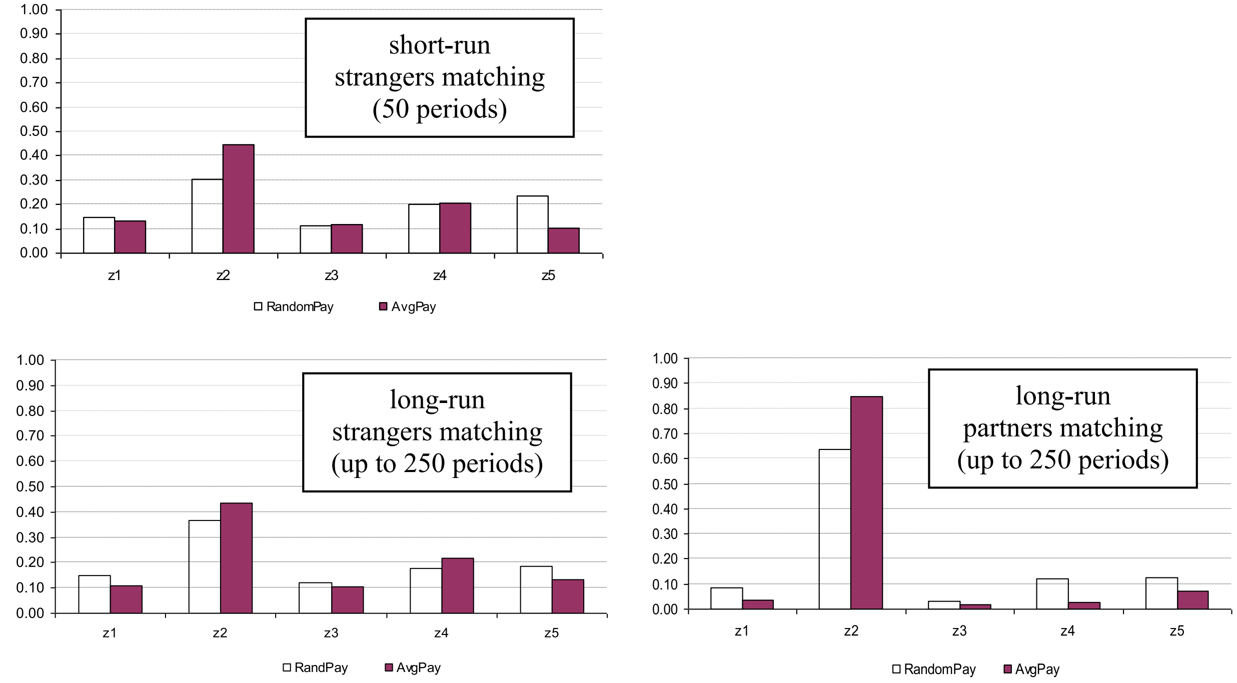


**Fig. 2.** Distribution of the game outcomes.

### 4.2. Overall distribution of strategy choices

Table 1 records the cumulative probabilities of playing $R$ for each player type. The overview of the observed strategies shows that the most frequent choices were pure strategies ($p_i = 0$ and $p_i = 1$).[5] Player 1 chose the imperfect equilibrium strategy $p_1 = 0$ and the perfect equilibrium $p_1 = 1$ with about equal frequencies (each about 20% of the time); player 2's most frequent choice was the non-equilibrium pure strategy $p_2 = 0$ (about 38% of the time); and player 3 chose to play the imperfect equilibrium strategy $p_3 = 1$ more frequently (about 30% of the time) than any other strategy.

Table 1 reveals only rather small differences in distribution of strategy choices comparing the payoff conditions in the strangers treatment. Comparing the distributions of the strategy choices in the short-run and the long-run strangers treatments in Table 1, we find that they are surprisingly similar across the different number of periods. The p-values of the Mann-Whitney *U test* comparing the observed relative frequencies in the short-run and long-run strangers treatments exceed the 10-percent level in all but one case.[6]

While the payoff conditions and the number of periods in the strangers treatment do not seem to affect the distribution of the strategy choices, the matching procedure does. The average share of pure strategies is 0.65 in the partners treatment, whereas it is 0.43 in the short-run and 0.37 in the long-run strangers treatments. The difference between the frequency of observed pure strategies across treatments is significant. Using the Mann-Whitney *U test*, we find p-values of 0.021 in the first and of 0.007 in the second comparison. Furthermore, we also observe a much higher frequency of pure strategy choices (more $p_1 = 0$ and $p_2 = 0$ as well as more $p_3 = 1$) in Average Pay than in the Random Pay condition of the partners treatment. This interaction effect between the treatment and the payoff conditions is the only manifest impact of the different payoff conditions.

***Observation 2a:*** In the strangers treatment, the distributions of strategy choices show no significant impact concerning the Average-Pay versus Random-Pay conditions or short-run versus long-run treatments.

***Observation 2b:*** In the partners treatment, the distributions of strategy choices show a significant treatment effect: The share of pure strategy choices is greater in Average Pay than in Random Pay and generally greater than in the strangers treatment.

Table 1 shows that the trembling-hand perfect equilibrium strategies ($p_1 = 1,\ p_2 = 1,\ 0 \le p_3 \le \frac{1}{4}$) are rarely chosen by the players. Unlike player 3's choices that tend to lean towards the imperfect equilibrium strategy, the distribution of choices by players 1 and 2

[5] Over all periods, rounds, and players, the relative frequency of pure strategies was 43%, with 24% $p_i = 0$ and 19% $p_i = 1$. The remaining 57% were mixed strategies.

[6] Only player 1's choice of playing right for sure, i.e., player 1's perfect equilibrium strategy $p_1 = 1$, differs (weakly) significantly between the short-run and long-run stranger's treatments (p-value of the Mann-Whitney test is 0.0935, two-sided). It is chosen less frequently in the long-run than in the short-run. This fact, however, does not seem to affect the relative frequencies of the outcome states that are statistically indistiguishable when comparing the short-run and the long-run stranger's macthing. We tested, comparing the short-run sessions periods to all periods of the long-run sessions, to the first 50 periods, and to the last 50 periods, separately.

**Table 1**
Cumulative distribution of subjects' behavior strategy.

| | | $p = 0$ | [0,.25] | [0,.33] | [0,.50] | [0,.75] | [0,.99] | [0,1] |
|---|---|---|---|---|---|---|---|---|
| **Short-run strangers treatment (50 periods)** | | | | | | | | |
| player 1 | AvgPay | **0.271** | 0.384 | 0.449 | 0.607 | 0.790 | 0.868 | 1.000 |
| | RandomPay | **0.247** | 0.312 | 0.330 | 0.470 | 0.629 | 0.801 | 1.000 |
| | Overall | **0.259** | 0.348** | 0.389 | 0.539 | 0.710 | 0.834 | 1.000* |
| player 2 | AvgPay | **0.443** | 0.671 | 0.719 | 0.836 | 0.878 | 0.936 | 1.000 |
| | RandomPay | **0.347** | 0.497 | 0.530 | 0.636 | 0.690 | 0.809 | 1.000 |
| | Overall | **0.395**** | 0.584** | 0.624 | 0.736 | 0.784 | 0.873 | 1.000 |
| player 3 | AvgPay | 0.061 | 0.128 | 0.147 | 0.270 | 0.443 | 0.616 | **1.000** |
| | RandomPay | 0.063 | 0.169 | 0.207 | 0.353 | 0.564 | **0.806** | 1.000 |
| | Overall | 0.062*** | 0.149 | 0.177 | 0.312 | 0.503 | 0.711 | **1.000** |
| **Long-run strangers treatment (up to 250 periods)** | | | | | | | | |
| player 1 | AvgPay | **0.192** | 0.356 | 0.407 | 0.563 | 0.756 | 0.892 | 1.000 |
| | RandomPay | **0.237** | 0.342 | 0.371 | 0.529 | 0.692 | 0.885 | 1.000 |
| | Overall | **0.229** | 0.361 | 0.398 | 0.542 | 0.728 | 0.890 | 1.000 |
| player 2 | AvgPay | **0.321** | 0.597 | 0.641 | 0.797 | 0.922 | 0.973 | 1.000 |
| | RandomPay | **0.294** | 0.467 | 0.496 | 0.682 | 0.808 | 0.882 | 1.000 |
| | Overall | **0.306** | 0.538 | 0.574 | 0.747 | 0.867 | 0.927 | 1.000*** |
| player 3 | AvgPay | 0.095 | 0.150 | 0.187 | 0.271 | 0.378 | 0.581 | **1.000** |
| | RandomPay | 0.075 | 0.168 | 0.191 | 0.304 | 0.506 | **0.785** | 1.000 |
| | Overall | 0.086 | 0.160 | 0.191 | 0.283 | 0.427* | 0.668 | **1.000** |
| **Long-run partners treatment (up to 250 periods)** | | | | | | | | |
| player 1 | AvgPay | **0.737** | 0.838 | 0.850 | 0.877 | 0.956 | 0.991 | 1.000 |
| | RandomPay | **0.431** | 0.590 | 0.623 | 0.860 | 0.878 | 0.900 | 1.000 |
| | Overall | **0.572** | 0.703 | 0.727 | 0.871* | 0.917 | 0.946 | 1.000 |
| player 2 | AvgPay | **0.564** | 0.600 | 0.610 | 0.663 | 0.712 | 0.867 | 1.000 |
| | RandomPay | **0.358** | 0.593 | 0.607 | 0.705 | 0.752 | 0.842 | 1.000 |
| | Overall | **0.461** | 0.604 | 0.615 | 0.687 | 0.731 | 0.855 | 1.000 |
| player 3 | AvgPay | 0.046 | 0.062 | 0.070 | 0.085 | 0.107 | 0.135 | **1.000** |
| | RandomPay | 0.100 | 0.127 | 0.143 | 0.177 | 0.234 | 0.547 | **1.000** |
| | Overall | 0.073 | 0.094 | 0.107 | 0.131** | 0.170 | 0.341* | **1.000***** |

**bold** numbers indicate the most frequent choice of the player type;.
*, **, *** (Mann-Whitney test result): non-cumulative relative frequency is significantly different between RandomPay and AvgPay at 10%, 5% and 1% level.

does not seem close to the trembling-hand perfect or imperfect equilibrium strategies. Although Player 1 chooses $p_1 = 0$ in about a quarter of the cases in the strangers treatment, the player's choices seem to span the entire spread between the imperfect and the perfect equilibrium strategies, with considerable probabilities in the entire range, i.e., $0 \leq p_1 \leq 1$. In the partners treatment, there seems to be a clear tendency for player 1 (about 50% of the choices) towards the imperfect equilibrium strategy, i.e., $p_1 = 0$.

The distribution of Player 2's choices is neither close to the perfect equilibrium strategy $p_2 = 1$ nor in the imperfect equilibrium strategy range, i.e., $\frac{1}{3} \leq p_2 \leq 1$. In contrast, we observe that about 50% to 60% of the choices by player 2, in all settings, involve probabilities smaller than 33%, i.e., smaller than in any equilibrium strategy. Instead, player 2's most frequent strategy choice is $p_2 = 0$. A closer look at the data shows this especially to be the case, when the other two players play according to their imperfect equilibrium strategies. We observe this particular type of strategy profile (0,0,1) in about one-quarter of the plays. These out-of-equilibrium choices are evidently not arbitrary, because a random draw from the entire strategy space would yield a much higher relative frequency (i.e., 2/3) of choices by player 2 that would be in line with the imperfect equilibrium strategy.

In about 5% of chosen strategies, players 1 and 2 actually make decisions that are simultaneously in line with their trembling-hand perfect equilibrium strategies. However, in most of these cases, player 3 deviates by choosing $p_3 = 1$ from the own trembling-hand perfect strategy $p_3 \leq \frac{1}{4}$. In fact, the perfect equilibrium strategy ($p_3 \leq \frac{1}{4}$) is rarely chosen by player 3 (on average <16% in any setting).[7] Since the observed frequency is even below the frequency expected by chance (26%), we conjecture that there is no support for deliberate perfect equilibrium play by player 3. Instead, in >30% of all cases in the strangers treatment and >60% of all cases in the partners treatment, player 3 plays the pure strategy $p_3 = 1$, which is in line with the imperfect equilibrium. Hence, the player 3 strategy choices seem more in line with the imperfect than with the trembling-hand perfect equilibrium.

Overall, we clearly have no evidence that the trembling-hand perfect equilibrium describes observed behavior in the Game of Selten's Horse. While we observe more instances of the imperfect than the trembling-hand perfect equilibrium outcome, the frequency of imperfect equilibrium plays is not high enough to conclude that it is the predominant outcome in the game.

***Observation 3:*** While the trembling-hand perfect equilibrium strategy profiles are less frequently observed in our data than the imperfect equilibrium strategy profiles, neither equilibrium concept convincingly describes observed aggregate behavior in the Game

[7] The highest relative frequency of the trembling-hand perfect equilibrium strategy by player 3 is 24% in one independent observation of the partners treatment, while the minimum is 1% in another.

of Selten's Horse.

### 4.3. Pareto efficiency and first mover advantage

The Game of Selten's Horse has a very interesting trait. Similar to social dilemma games, the trembling-hand perfect equilibrium deviates from Pareto efficiency. A necessary condition to reach a Pareto efficient allocation in the Game of Selten's Horse (i.e., a state in which no player can get better off without making another player worse off) is that player 3 chooses $R$ for sure (i.e., $p_3 = 1$). The imperfect equilibrium outcome $z_2$, the non-equilibrium outcome $z_4$, and all mixed outcomes between $z_2$ and $z_4$ are Pareto efficient. We observe that more than half of the outcomes in our experiment are Pareto efficient.[8]

Note, however, that the Pareto efficient outcomes (i.e., when player 3 plays $p_3 = 1$) can only be sustained, if player 1 plays $L$ with positive probability to ensure that the game does not consistently end in $z_4$, where player 3 receives zero. If player 3′s expected payoff from playing pure strategy $R$ is below 1, she will reduce the probability of playing $R$ to increase her expected payoff.

Note also that the game outcome crucially depends on the strategy choice of player 1, who seems to have a special kind of first-mover advantage in the Game of Selten's Horse. By choosing any strategy $p_1 \in [0,\ 0.5]$, player 1 can enforce a Pareto efficient outcome that is a mixture of $z_2$ and $z_4$. While $p_1 \leq 0.5$ is not an equilibrium strategy (i.e., it is not the best response to the other players' strategies $p_2 = 0$ and $p_3 = 1$), it is an attractive strategy for player 1 for two reasons. First, it leads to a greater expected payoff than player 1 can achieve in any equilibrium of the game. Second, $p_1 \leq 0.5$ is sustainable, because player 2 and player 3 have no reason to deviate from their best responses $p_2 = 0$ and $p_3 = 1$.

As can be seen in Table 1, the majority of the strategy choices by player 1 in the strangers treatment and an even greater share in the partners treatment are in the interval $p_1 \in [0, 0.5]$. At the same time, we find that the modal choices of players 2 and 3 confirm the indicated pure best response strategies ($p_2 = 0$ and $p_3 = 1$) to player 1′s strategy choice described above, i.e., $p_1 \in [0,0.5]$.[9] Especially, in the long-run partners treatment, we observe that close to half of the strategy choices by player 2 are $p_2 = 0$ and almost two-thirds of those by player 3 are $p_3 = 1$. Overall, these Pareto efficient strategy combinations are the most frequently observed choices.

***Observation 4:*** While most Pareto efficient strategy combinations $\{p_1 \in (0,\ 0.5),\ p_2 = 0, p_3 = 1)$ are not equilibria of the Game of Selten's Horse, they are the most frequently observed choices, with a substantially and significantly higher relative frequency in the partners treatment than in the strangers treatment.

## 5. Behavioral dynamics

After having discussed the distributions of decision and outcomes in the last section, we take a closer look at the dynamics of the observed behavior in this section. The underlying question is whether there are behavioral dynamics in the Game of Selten's Horse that drive decisions and outcomes towards one of the discussed equilibria or towards some other rest point. A first glance at the data suggests that subjects adjust their strategies in an ex-post best response manner. When player 1 increases the probability of playing $R$, player 3 frequently decreases the probability of choosing $R$, and vice versa. When player 3 increases the probability of playing $R$, player 2 frequently decreases the probability of choosing $R$, and vice versa. In the following, we investigate the behavioral dynamics more closely.

### 5.1. Learning direction theory

First, we apply learning direction theory (Selten and Stoecker 1986; Selten and Buchta 1999) to the data. According to learning direction theory, players adjust their choices in hindsight, either by moving them towards the ex-post best responses or by leaving them unchanged. Selten and Buchta (1999) illustrate learning direction theory using the analogy of an (autodidactic) marksman who learns how to hit a trunk with a bow and arrow:

"If he misses the trunk to the right, he will shift the position of the bow to the left and if he misses the trunk to the left, he will shift the position of the bow to the right. The marksman looks at his experience from the last trial and adjusts his behavior." (p. 86, Selten and Buchta 1999).

The ex-post best response function (2) takes the other players' actions as given and determines the best response to these given actions.[10] We can use the best response functions given in Section 2. In each best response function, we replace the strategy choices of the other players by the last period's relative frequency, $f_{-i}^{t-1}$, of choosing $R$:

[8] We find a somewhat lower frequency of Pareto efficient outcomes in the Random Pay condition than the Average Pay condition. However, the difference is not significant. Given the similarity of the outcomes, hereafter, we report only on the pooled data without distinguishing between Average Pay and Random Pay any longer.

[9] Particularly, player 2 deviates from playing the equilibrium strategy. In any equilibrium, player 2 should play $R$ with a probability of at least 0.25. The data suggest that the majority of player 2 choices violate that prediction. The triggering point of such behavior is that player 3 chooses $R$ with a probability above 0.25, thus shifting player 2's best response from $R$ to $L$.

[10] Note the similarity to the Cournot learning model, which suggests that the ex-post best response should be chosen in the following period. Learning direction theory postulates that adaptation of the strategy in the direction of the ex-post best response is more likely than adaptation in the opposite direction, however, without specifying the magnitude of the strategy adjustment.

$$\begin{aligned} p_{1t} &= \begin{cases} 1 & \\ [0,1] & if\ 4(1-f_2^{t-1})f_3^{t-1} + f_3^{t-1} \begin{matrix} < \\ = \\ > \end{matrix} 3f_3^{t-1} \\ 0 & \end{cases} \\ p_{2t} &= \begin{cases} 1 & \\ [0,1] & if\ f_3^{t-1} \qquad\qquad\qquad\quad \begin{matrix} < \\ = \\ > \end{matrix} 0.25 \\ 0 & \end{cases} \\ p_{3t} &= \begin{cases} 1 & \\ [0,1] & if\ 2(1-f_1^{t-1}) \qquad\qquad \begin{matrix} > \\ = \\ < \end{matrix} f_1^{t-1}(1-f_2^{t-1}) \\ 0 & \end{cases} \end{aligned} \tag{2}$$

The feedback information on the distribution of outcomes allows subjects to (approximately) infer the strategies of the other two players. According to learning direction theory, subjects are more likely to adapt their strategy in the direction of their ex-post best response than in any other direction. This impulse to adapt the strategy is obviously absent if the subject played the best response in the previous period. In that case, learning direction theory predicts no change.

We can test learning direction theory on the individual or the group level. On individual level, we measure whether subjects exhibit more changes in the predicted than in the opposite direction. The mid-section of Table 2 reports the number and the percentage of subjects in each setting, who made more changes in the predicted than in the opposite direction. In all settings and for all but one player type (player 3 in the long-run partners treatment), the great majority of subjects exhibit a higher frequency of adapting their strategy choice from one period to the next in the direction suggested by learning direction theory than in the opposite direction.

Instead of only counting the cases in which a subject makes a predicted or unpredicted change, we can additionally also count the cases in which subjects repeat their last choice, i.e., make no change. These cases are – strictly speaking – also in line with learning direction theory. Table 2, in the right-hand section, reports the number and the percentage of subjects, who either adapt their choices in the predicted direction or keep them unchanged. Using this more inclusive definition, we find that an overwhelming majority of subjects adhere to the behavioral adaptation suggested by the learning direction theory.

In fact, we observe that the adherence to the learning direction theory is not only true on an individual level for the majority of all subjects, but also true for the majority of subjects in almost every independent observation. Hence, we can conclude that learning direction theory governs the behavioral dynamics of almost all subjects and almost all independent observations.[11] Furthermore, it seems to organize all player types almost equally. The only exception seems to be player 3 in the partners treatment. The subjects of that type exhibit a lower degree of adaptation and a higher degree of inertia than all other subjects. Nevertheless, their behavior is in line with learning direction theory for about two-thirds of the subjects.

***Observation 5:*** Subjects' behavior is in line with learning direction theory.

### 5.2. *Impulse response and impulse balance*

Impulse balance theory describes the long-term attraction point of the dynamics of learning direction theory (Selten 2004; Selten, Abbink and Cox 2005; Ockenfels and Selten 2005; Neugebauer and Selten 2006). Ex-post rationality results in a positive or negative impulse vis-à-vis the pure strategy $R$ in accordance with learning direction theory. If the dynamics come to rest, we have an impulse balance point where positive and negative impulses cancel out.[12]

In the Game of Selten's Horse, impulse balance points can be determined by the rest points of the *impulse response trajectories*, which result from an adaptive simulation procedure closely related to Chmura, Goerg and Selten (2012).[13] Accordingly, the probabilities of playing $R$ in the Game of Selten's Horse are updated after each round of feedback taking account of the most recently received impulses. A positive impulse affects a movement of player $i$'s strategy by one step in the direction of $R$ in agreement with the best response dynamics (3). The step length in our case is $r_i(t) = .01$. A negative impulse affects a corresponding decrease of the *simulated behavior strategy* $\widetilde{p}_{it}$ by one step. If no impulse is given, e.g., in the impulse balance point, the adjustment process rests.

[11] Compared to studies of less complex games we find lower agreement with learning direction theory. Neugebauer and Selten (2006), e.g., report that only 8% of their subjects' choices were at odds with learning direction theory.

[12] In most specifications, the impulses from losses are weighted more strongly than those from gains (see Selten and Chmura 2008, Selten, Chmura and Goerg 2011, Chmura, Goerg, and Selten 2012). Selten, Abbink and Cox (2005) argue that differential weighting is in line with loss aversion. In the Game of Selten's Horse, however, differential weighting is not necessary (and not used), because all payoffs are in the domain of gains, compared to the maximin outcome of 0.

[13] The difference between impulse response dynamics and impulse matching dynamics lies in the updating rule and the step-length. While the former is deterministic and only considers the impulses resulting from one-period hindsight and applies one-step adjustment, the latter is stochastic and adds up all previous periods' impulses to create long-term drivers for upwards versus downwards adaptation of behavior. Goerg, Neugebauer and Sadrieh (2016) apply our concept of impulse response dynamics to the minimum effort game. Chmura and Güth (2011) investigate impulse matching dynamics in the minority game.

**Table 2**
Evidence in favor of learning direction theory.

| Treatment | Number of subjects per player type | Number of subjects with more changes in the predicted than in the opposite direction | | | Number of subjects whose responses are in line rather than at odds with learning direction theory | | |
|---|---|---|---|---|---|---|---|
| | | Player 1 | Player 2 | Player 3 | Player 1 | Player 2 | Player 3 |
| short-run strangers | 54 | 38 | 40 | 41 | 47 | 47 | 47 |
| | 100% | 70.4% | 74.1% | 75.9% | 87.0% | 87.0% | 87.0% |
| long-run strangers | 54 | 42 | 40 | 51 | 45 | 43 | 52 |
| | 100% | 77.8% | 74.1% | 94.4% | 83.3% | 79.6% | 96.3% |
| long-run partners | 18 | 14 | 13 | 6 | 18 | 13 | 12 |
| | 100% | 77.8% | 72.2% | 33.3% | 100% | 72.2% | 66.7% |

$$\widetilde{p}_{it} = \widetilde{p}_{it-1} + r_i(t),\ \textit{where}$$
$$r_i(t) = \begin{cases} .01, & \textit{if} \quad p_{it} = 1 \wedge \widetilde{p}_{it-1} < 1, \\ -.01, & \textit{if} \quad p_{it} = 0 \wedge \widetilde{p}_{it-1} > 0, \\ 0 & \qquad \textit{otherwise}. \end{cases} \tag{3}$$

For a given initial strategy profile (0.5, 0.5, 0.5), Fig. 3 exhibits the impulse response trajectories of the three players towards the impulse balance of the game (1, 1, 0). In fact, the thus encountered impulse balance point equals the trembling-hand perfect pure strategy equilibrium in the Game of Selten's Horse. This finding is curious. On the one hand, our data do not show much support of the trembling-hand perfect equilibrium. On the other hand, however, we find that the dynamics of the impulse balance that are based on direction learning behavior, which we observed for most of our subjects, should lead to the trembling-hand perfect equilibrium.

We take a closer look at the dynamics in the data by studying the impulse response trajectories for each independent observation separately. Our simulations ignore the feedback actually received by subjects, instead replaying the original random matches of interacting subjects after the first period and using the simulated choices for feedback. The simulation is deterministic as it uses the original matches and the original starting point. If the subjects behaved in accordance with the one-step impulse response adjustment, the simulated trajectories would be identical to the observed ones. The figures A3.1 to A3.6 in the appendix A3 present the resulting simulations in a separate chart for each independent observation together with the smoothed trajectories of the observed choices. The observed behavior strategies of each player type and their simulations are averaged over ten periods for each independent observation. Each mark in the chart represents the ten-period average of the decisions of the subjects for each player type.

For most groups in the short-run strangers treatment the simulated impulse response trajectories are impressively close to the observed behavioral trajectories (Figures A3.1 and A3.2). Note, however, that while all simulated impulse response trajectories converge on the impulse balance point (i.e., the trembling-hand perfect equilibrium), observed behavior seems to need more than 50 periods to converge. By the end of the short-run sessions (i.e., at $t = 50$), the trajectory for player 3 seems to be still moving towards $R$ and the trajectory for player 2 towards $L$, corresponding to the very first part of the impulse response trajectories in Fig. 3.

Prima facie, it may seem that giving subjects more repetitions in a session would lead to full convergence of behavior on the perfect equilibrium. Since subjects' choice trajectories in the 50 periods seemed close to the impulse balance trajectories, giving subjects more periods than the 165 periods needed for convergence in the simulation (see Fig. 3) should potentially result in a high frequency of perfect equilibrium outcomes. However, the observed trajectories of choices in the long-run strangers treatment (Figures A3.3 and A3.4) show otherwise. The overall impression in the long-run strangers treatment is that the impulse response trajectories rarely follow the observed choices over the long-run time horizon. A typical long-run choice trajectory starts at an intermediate probability of playing $R$, then moves either to a higher or a smaller non-extreme probability, where it then remains relatively constant until the end of the session. The simulated impulse response trajectories, in contrast, follow their complex paths towards the perfect equilibrium. It seems that player 1 generally balances $R$ and $L$ choices, while player 2 usually choose to play $L$ rather than $R$ and player 3 plays $R$ rather than $L$. This kind of non-equilibrium behavior makes convergence on any equilibrium unlikely.[14]

Comparing the convergence patterns between the long-run strangers and long-run partners treatments, we find substantial differences for players 1 and 3, but not for player 2. Fig. 4 displays the observed trajectories of play in the first 50 and the last 50 periods of the long-run sessions with strangers and partners treatments. Player 2′s strategy choices are very similar in both the partners and the strangers treatments, with the probability of playing $R$ dropping slightly early on, but then remaining close to the lower bound of the imperfect equilibrium strategy, in the range between 30% and 40%, for the rest of the periods.

[14] If anything, the observed strategy adjustments in the long-run strangers' treatment indicate a move away rather than towards the trembling-hand perfect equilibrium. In the first 50 periods, the relative frequency of the trembling-hand perfect-equilibrium outcome ($z_5$) is 0.1887, but it is only 0.1522 in the last 50 periods. The difference is weakly significant as the two-tailed Wilcoxon signed ranks test confirms with a p-value of 0.0936.

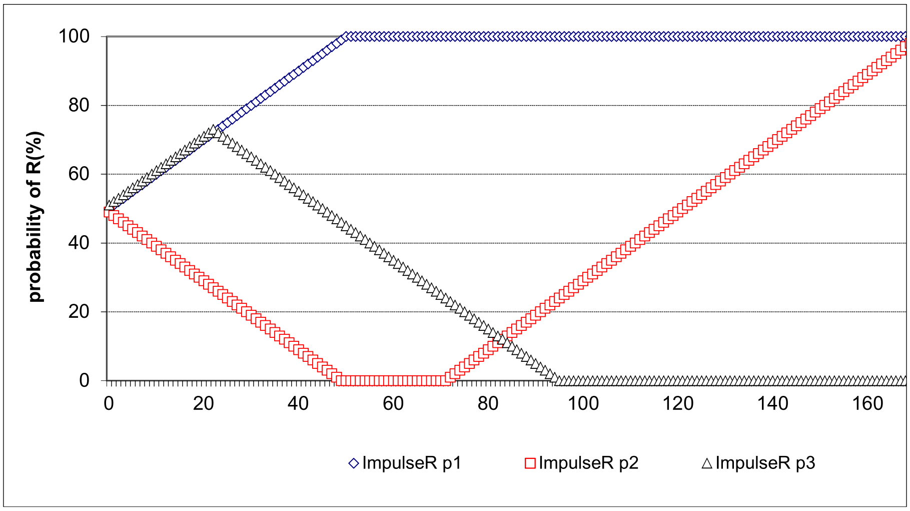


**Fig. 3.** Impulse response trajectory of strategies chosen by players of type 1 (filled square), 2 (empty diamond), 3 (empty triangle) with initial profile (0.5, 0.5, 0.5).

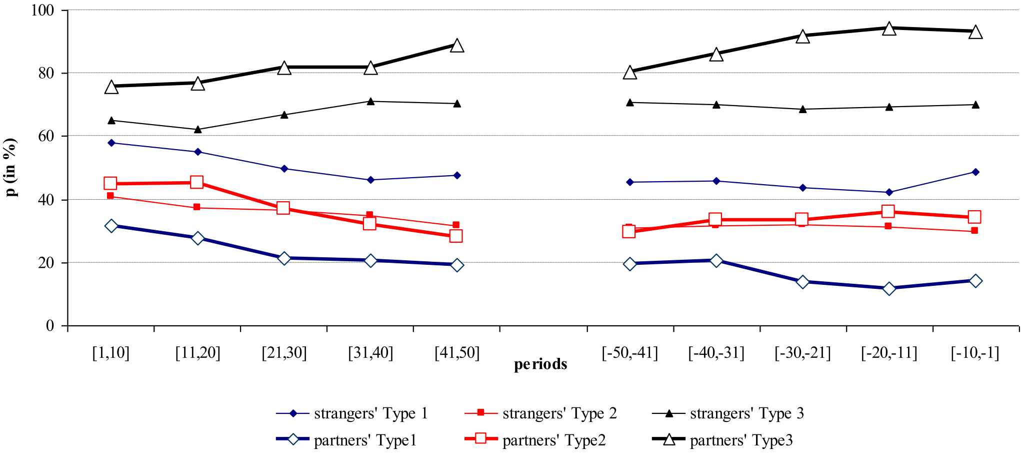


**Fig. 4.** Average behavior strategies of players 1, 2, and 3 in the long-run experiments.

Unlike the strategy choices by player 2, those by players 1 and 3 are clearly different between the partners treatment and the strangers treatment. In the long-run strangers treatment, we observe a majority of the player 1 and player 3 choices well out-of-equilibrium without an apparent tendency to converge, with $p_1$ hovering around 50% and $p_3$ around 70%. In contrast, the trajectories of observed play in the long-run partners treatment seem to converge on the imperfect equilibrium strategies, with $p_1$ approaching 0% and $p_3$ approaching 100%. Hence, in the partners treatment the behavior strategies of all three player types are close to their imperfect equilibrium play. That explains why the frequency of the imperfect equilibrium outcome $z_2$ is much higher and the frequency of the non-equilibrium outcome $z_4$ lower in the partners treatment than in the strangers treatment (see Fig. 2).

Each mark represents the ten-period average probability of playing *R* in the first and the last fifty periods. Thin lines/filled dots (thick lines/empty dots) show the trajectories of the strangers (partners) treatment.

Although direction learning is in line with the choices of the majority of subjects in both long-run strangers treatment and long-run partners treatment (see Table 2), the impulse response trajectories (see Figures A3.3 through A3.6 in the appendix) seem to provide no good fit for the long-run behavior in the Game of Selten’s Horse. Impulse response trajectories converge on the trembling-hand perfect equilibrium in the long-run, but the observed long-run behavior does not. While the impluse response dynamics seem in line with the observed short-run behavior, in the long run, observed trajectories seem to move towards the Pareto efficient outcomes. Repeated group interaction is frequently quite stable over time, quickly converging on a Pareto efficient outcome, which in the long-run partners treatment is often the imperfect equlibrium.

***Observation 6a:*** In the short-run strangers treatment, the observed trajectories of choices are close to the impulse response trajectories.

***Observation 6b:*** In the long-run strangers treatment, the observed trajectories of choices are close to the impulse response trajectories early on, but do not converge on the perfect equilibrium as predicted by the impulse response dynamics. Instead, behavior often settles in the range of Pareto efficient outcomes with player 1 choosing $R$ in about 50% of the plays.

***Observation 6c:*** In the long-run partners treatment, the observed trajectories of choices are close to the impulse response trajectories early on, but do not converge on the perfect equilibrium as predicted by impulse response dynamics. Instead, behavior often moves towards the imperfect equilibrium that is in the range of Pareto-efficient outcomes.

### *5.3. Reinforcement learning*

Reinforcement learning (Erev and Roth 1998) offers an alternative concept for the dynamics of play that has been shown to capture the dynamics of behavior well in numerous game settings. As with the impulse response simulations, we start the reinforcement learning simulations at the observed initial choices in each independent group and replay the original random matches. The reinforcement learning dynamics (4) are as follows:

$$\begin{aligned}
&\widetilde{p}_{it} = \frac{\alpha^R_{it}}{\alpha^L_{it} + \alpha^R_{it}} \text{ , where} \\
&\alpha^R_{it} = \alpha^R_{it-1} + e^R_{it-1}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1}) \text{ ,} \\
&\alpha^L_{it} = \alpha^L_{it-1} + e^L_{it-1}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1})
\end{aligned} \tag{4}$$

with $e^j_{it-1}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1})$ denoting the expected payoff when playing $j = L, R$ conditionally of the other players $-i$ employing their simulated behavior strategy, and $\alpha^R_{i1}$ being the observed first behavior strategy of subject $i$ multiplied by 100. We measure the mean squared deviation of each subject's reinforcement trajectory from the observed trajectory over each 10-period interval and sum the deviations over the players and the time intervals. The simulation is deterministic as the original random matches for each period of the independent observations are replayed and the expected payoffs are used for probability adjustment.[15] Comparing the mean squared deviations of the reinforcement learning trajectories to the mean squared deviations of the impulse response trajectories in Table A2.1 (see appendix A2), we find that in 11 of 18 sessions in the short-run strangers treatment, in 9 of 18 sessions in the long-run strangers treatment, and in 3 of the 18 sessions in the long-run partners treatment, the latter is smaller than the former. Hence, reinforcement learning predicts the outcomes of the choices neither much better nor much worse than impulse response dynamics. Impulse response seems to be slightly better than reinforcement learning in the short-run strangers treatment and reinforcement learning clearly better than impulse response in the long-run partners treatment.[16]

We note that the reinforcement dynamics do not necessarily converge on an equilibrium. In contrast to the non-monotonic adjustment dynamics of the impulse response model towards the perfect equilibrium, the reinforcement trajectories are monotonic in the Game of Selten's Horse. Each reinforcement trajectory converges towards the upper or lower boundary of the behavior strategies. Once a boundary is reached by a reinforcement trajectory, the trajectory settles there. Typically, the reinforcement trajectories move fast in early periods, slowing down more and more as the simulation proceeds. Hence, simulations need a very long time (many thousands of periods) to converge on the rest point. Depending on the starting point different pure strategy profiles can institute an attraction point. Hence, reinforcement dynamics do not necessarily converge on an equilibrium, even though they do seem to capture some of the behavioral dynamics, especially early on in the game.

#### *5.3.1. Reinforcement one-step version*

For the ease of comparison with the impulse response trajectories and to serve as a basis for a mixed model presented in the following subsection, we propose a one-step adjustment version of reinforcement. The reinforcement one-step adaptation model is described as follows:

[15] We also conducted reinforcement learning simulations with binary stochastic choice. Since the simulated payoffs change with the binary choice, each simulation run leads to different trajectories. Therefore, we conducted 1,000 simulation runs for each independent observation. The average squared deviations across these simulations are large, much larger than the deviations of the reported deterministic simulations. So, we report the results concerning the latter one only.

[16] Using the Wilcoxon signed-ranks test, we only find a significant difference between the two predictions for the case of the long-run partners' treatment, where reinforcement learning dynamics seem to fare better than the impluse balance dynamics (p = .011, two-tailed).

$$\begin{aligned}
&\widetilde{p}_{it} = \widetilde{p}_{it-1} + s_i(t),\ \textit{where}\\
&s_i(t) = \begin{cases} .01, & \textit{if}\quad \widetilde{p}_{it-1} < \pi_{it},\\ -.01, & \textit{if}\quad \widetilde{p}_{it-1} > \pi_{it},\\ 0, & \quad\textit{otherwise}\end{cases}\\
&\pi_{it} = \frac{\alpha^R_{it}}{\alpha^L_{it} + \alpha^R_{it}}\\
&\alpha^R_{it} = \alpha^R_{it-1} + e^R_{it-1}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1}),\\
&\alpha^L_{it} = \alpha^L_{it-1} + e^L_{it-1}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1})
\end{aligned} \tag{5}$$

This simulation is also deterministic. Different from the model in the previous subsection, simulated choices involve only the possible mixed strategy choices, $\widetilde{p}_{it} = \{.00,.01,.02, \ldots,1.00\}$. The corresponding mean squared deviations of the one-step reinforcement learning trajectories to the observed choices are reported in Table A2.1 (see appendix A2). We find that in 11 of 18 sessions in the short-run strangers treatment, in 10 of 18 sessions in the long-run strangers treatment, and in 6 of the 18 sessions in the long-run partners treatment, the deviation is smaller for the impulse response than for the one-step reinforcement model.

***Observation 7:*** Reinforcement learning trajectories are closer to the observed trajectories than impulse response trajectories in the long-run partners treatment, but not in the short-run and long-run strangers treatment.

### 5.3.2. *Mixed learning model*[17]

The literature suggests that mixture models involving reinforcement and belief learning could be particularly well-suited for capturing the dynamics of interactive play (Camerer, 2003). In line with this perspective, we evaluate the predictive success of a mixed model simulation that incorporates impulse response and one-step reinforcement learning. Putting Eqs. (3) and (5) together, we propose the following mixed learning model.

$$\begin{aligned}
&\widetilde{p}_{it} = \widetilde{p}_{it-1} + \tau_i(t),\ \textit{where}\\
&\tau_i(t) = \begin{cases} .01, & \textit{if}\quad r_i(t) + s_i(t) > 0,\\ -.01, & \textit{if}\quad r_i(t) + s_i(t) < 0,\\ 0, & \quad\textit{otherwise}.\end{cases}
\end{aligned} \tag{6}$$

The simulation model is deterministic and follows the same procedures as before. It uses the observed initial choices as starting points and replays the original matches of the experiment. The approach is parameter-free, utilizes feedback from the simulated choices, and permits the same mixed strategy choices that were used in the experiment, $\widetilde{p}_{it} = \{.00,.01,.02, \ldots,1.00\}$.

We also report the corresponding mean squared deviations of the mixed learning model trajectories to the observed choices in Table A2.1 (see appendix A2). The results suggest that in 11 of 18 sessions in the short-run strangers treatment, in 9 of 18 sessions in the long-run strangers treatment, and in 5 of the 18 sessions in the long-run partners treatment, the impulse response implies a smaller deviation from the observed choices than the mixed model. Hence, we confirm observation 7 also for the mixed learning model.

## 5.4. *Quantal response equilibrium*

Similar to the concept of trembling-hand perfection, the quantal response approach (McKelvey and Palfrey 1995) allows for players to make errors. Particularly, initial choices of inexperienced players are assumed to be noisy in the quantal response approach, assigning an equal probability to each strategy. However, while trembling-hand perfection selects the equilibrium based on its robustness against errors, quantal response selects the strategy profile that is reached as errors eventually vanish.

The following set of equations shows the logit quantal-response functions (7) for the Game of Selten's Horse with the noise parameter $\lambda^{-1}$, where $\lambda$ is assumed to be close to zero for inexperienced players and large for experienced subjects.

$$\begin{aligned}
p_1(p_2, p_3, \lambda) &= \frac{1}{1 + \exp(-\lambda(p_2 + p_3 - 4p_2p_3))}\\
p_2(p_1, p_3, \lambda) &= \frac{1}{1 + \exp(-\lambda p_1(1 - 4p_3))}\\
p_3(p_1, p_2, \lambda) &= \frac{1}{1 + \exp(-\lambda(2 - 3p_1 + 2p_1p_2))}
\end{aligned} \tag{7}$$

With perfect noise, i.e., $\lambda = 0$, the proposed strategy profile is (0.5, 0.5, 0.5) and when noise vanishes, i.e., as $\lambda \to \infty$, the quantal

[17] Our adaptive approach of impulse response is outcome oriented and is parameter free. Hence, a comparison of the impulse response dynamics with reinforcement dynamics is straight forward. Belief learning models may also apply to our data (e.g., Cheung and Friedman 1997, Nyarko and Schotter 2002), or hybrid models as, in particular, the experienced weighted attraction model (Camerer 2002, Ho, Camerer and Chong 2007). However, for our setting these models require additional assumptions as beliefs are unobservable.

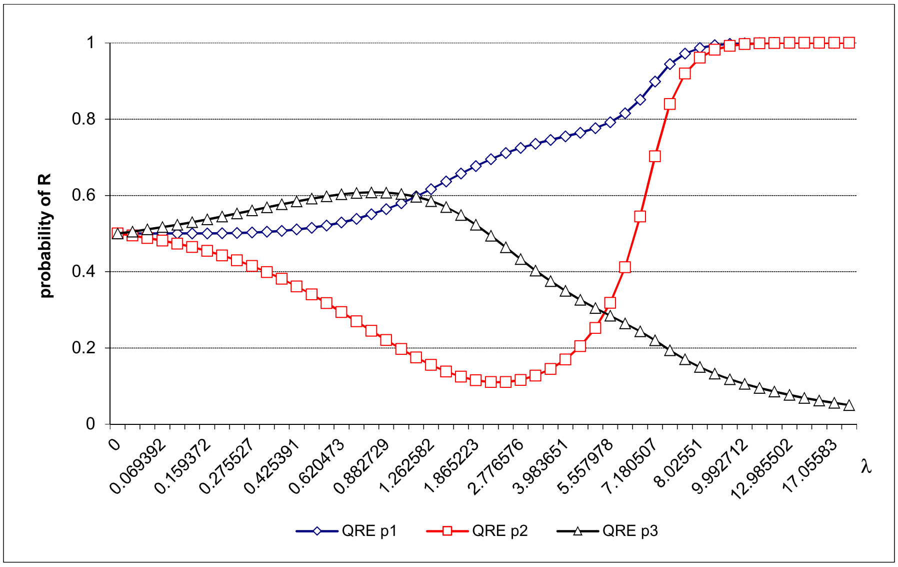


**Fig. 5.** Principal branch of quantal response correspondence.

response correspondence approaches the set of perfect equilibrium profiles ($p_1 = 1$, $p_2 = 1$, $p_3 \leq 0.25$). Fig. 5 displays the quantal response curves, which are attracted to the trembling-hand perfect equilibrium.[18] We note the similarity of Fig. 5 to Fig. 3. The quantal response curves in Fig. 5 look like a smooth version of the impulse response trajectories in Fig. 3. Also note that before quantal response reaches the perfect equilibrium set at $\lambda \geq 18$, predicted play in the quantal response equilibria corresponds to non-equilibrium behavior in a setting with fully rational players.

The initial choices of subject types 1 and 2 in our experiment are almost uniformly distributed over the interval [0,1] with modal choices 0 and 0.5, and the initial choices of subject type 3 over [0.4,1] with modal choices 0.5 and 1.

Similarly to Capra et al. (1999) and Goeree and Holt (1999), we estimate $\lambda = 0.31$ (s.d. $= 0.029$), applying the MLE to the overall data in the short-run strangers treatment. For the first period, we have an estimate of 0.00 (s.d. $= 0.100$). For the first and the last ten periods, our estimates of $\lambda$ are 0.03 (s.d. $= 0.031$) and 0.45 (s.d. $= 0.028$), correspondingly. These estimates indicate a reduction of noise and a movement towards the trembling-hand perfect equilibrium over time in the short-run strangers treatment.

Applying the logit quantal response model to the overall data in the long-run strangers treatment, we obtain the estimate $\lambda = 1.00$ (s.d. $= 0.13$). Our estimate for the first period data is $\lambda = 0.00$ (s.d. $= 0.13$). For the data of the first and last ten periods, we estimate $\lambda = 0.19$ (s.d. $= 0.11$) and $\lambda = 1.07$ (s.d. $= 0.14$), correspondingly. Hence, the estimated noise level of choices in the long-run strangers treatment seems to start similarly but decrease more as compared to the short-run strangers treatment. Obviously, the larger number of periods can contribute to more experience and less confusion in the game, leading to decreased noise levels, particularly in later periods towards the end of the sessions.

For the long-run partners treatment, we obtain the following estimates of the logit quantal response model. We estimate $\lambda = 1.22$ (s. d. $= 0.24$), when applying the MLE to the overall data. For the first and last 10 periods, our estimates are 1.65 (s.d. $= 0.25$) and 1.17 (s. d. $= 0.36$), correspondingly. It is not surprising that playing in stable partners matchings leads to lower noise levels than playing in the randomly rematched groups of the strangers treatment. Note, however, that in the long-run partners treatment we do not observe decreasing noise levels over time as we do in the short-run and the long-run strangers treatment and as would be in line with the learning dynamics in a quantal response model. In fact, the noise level estimate for the last ten periods is even higher (i.e., the $\lambda$ value lower) than for the first ten periods.

Overall, our estimated quantal response models show $\lambda$ values that are all in the range of about 1 to 2 and, thus, far below 18, the threshold for $\lambda$ values that are necessary for convergence on the perfect equilibrium point. In other words, even after many periods of play, the noise levels in the estimated quantal response models remain very high in all treatments. The persistent prevalence noisy choice behavior throughout all periods seems to indicate that behavior in the Game of Selten's Horse does not follow the trajectories suggested by quantal response dynamics that converge on the perfect equilibrium (see Fig. 5).

***Observation 8:*** The principal branch of the quantal response correspondence converges on the trembling-hand perfect equilibrium

[18] We made use of the Gambit software (McKelvey, McLennan, Turocy2006) to compute the principal branch of the logit quantal response correspondence.

**Table 3**
Level-k responses in the Game of Selten's Horse.

| | behavior strategies | | | cycle steps | outcomes without k-level-mixtures | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Level k | **p1** | **p2** | **P3** | | z1 | z2 | z3 | z4 | z5 |
| 0 | 0.5 | 0.5 | 0.5 | – | 0.25 | 0.25 | 0.125 | 0.125 | 0.25 |
| 1 | 0.5 | 0 | 1 | – | 0 | 0.5 | 0 | 0.5 | 0 |
| 2 | 1 | 0 | 1 | step 1 | 0 | 0 | 0 | 1 | 0 |
| 3 | 1 | 0 | 0 | step 2 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0.5 | 1 | 0 | step 3 | 0.5 | 0 | 0 | 0 | 0.5 |
| 5 | 1 | 1 | 1 | step 4 | 0 | 0 | 0 | 0 | 1 |
| 6 | 0 | 0 | 0.5 | step 5 | 0.5 | 0.5 | 0 | 0 | 0 |
| 7 | 1 | 0 | 1 | step 1 | 0 | 0 | 0 | 1 | 0 |
| 8 | 1 | 0 | 0 | step 2 | 0 | 0 | 1 | 0 | 0 |
| 9 | 0.5 | 1 | 0 | step 3 | 0.5 | 0 | 0 | 0 | 0.5 |
| 10 | 1 | 1 | 1 | step 4 | 0 | 0 | 0 | 0 | 1 |
| 11 | 0 | 0 | 0.5 | step 5 | 0.5 | 0.5 | 0 | 0 | 0 |
| 12 | 1 | 0 | 1 | step 1 | 0 | 0 | 0 | 1 | 0 |
| … | … | … | … | … | … | … | … | … | … |

in the Game of Selten's Horse when noise vanishes. The observed trajectories seem not in line with this behavioral prediction, especially because the estimated noise levels remain very high even after many periods in all treatments.

### 5.5. *Level-k non-equilibrium model*

The level-k model of cognitive reasoning (Nagel 1995; Stahl and Wilson 1995) implies a hierarchy of best-response modes. The standard approach of Crawford (2013) assumes that players with no strategic reasoning (i.e., level-0 types) make random choices. Level-1 types play best response to level-0 types, level-2 types play best response to level-1 types, and so forth. Generally, level-k type players play best response to level-(k-1) type players. In many games, as for instance the centipede game (García-Pola et al., 2020), level-k reasoning converges to common knowledge of rationality as $k \to \infty$. In the Game of Selten's Horse, however, level-k responses do not converge with an increasing level of reasoning. Instead, they cycle. Table 3 shows the level-k responses for the levels $k = \{0, 1, \ldots, 12\}$. The first cycle starts at level-2 and ends after five steps at level-6. Then, the next cycle begins at level-7 and ends at level-11, again after five steps that are identical to those in the first cycle. These cycles are then repeated over and over again without any variation or convergence.

Due to the cycles and due to the fact that only three different types of responses are contained in the level-k responses (i.e., $p_i = 0$, $p_i = 0.5$, or $p_i = 1$), there is no one-to-one mapping of observed behavior to the strategic reasoning level of a subject. For example, observing the behavior strategy $p_1 = 1$ may correspond to a player 1 who is reasoning on any level-k, where k ϵ {2, 3, 5, 7, 8, 10, 12, …}. This holds similarly true for any of the other two behavior strategies and players.

Since the identification of the level of reasoning on an individual level is not possible, we employed a population mixture identification strategy for our level-k analysis. Using a least-squares method, we identify the mixture of level-0, level-1, level-2, and level-3 players that induces a behavior strategy profile ($p_1$, $p_2$, $p_3$) that is closest to the one we observe in our data. Table 4 shows the level-k mixtures that we identify overall and for each of the treatments.[19]

All in all, our level-k analysis seems to indicate that the level of reasoning used by subjects in the Game of Selten's Horse is not very high. Substantially more than 50 percent of the subjects are identified as playing level-0 and the others reveal only a level-1 reasoning. However, as discussed above in Section 4.3, we do not believe that the observed behavior strategies are due to low levels of strategic reasoning, but due to the specific structure of the game, in which player 1 can successfully enforce Pareto efficient outcomes by choosing a behavior strategy close to the 50–50 mixture (i.e., $p_1 = 0.5$). This form of enforced cooperation probably entails a high level of reasoning, even though it is identified as level-0 behavior in the level-k model.

***Observation 9:*** The Level-k model does not converge with increasing cognitive reasoning in the Game of Selten's Horse, but cycles instead. The estimated levels of reasoning in the observed data are extremely low with >60% on level zero.

## 6. Conclusions

We report experimental data on the Game of Selten's Horse (Selten 1975), which is a 3-person game with two pure strategies (*R*ight or *L*eft) for each player. The game was designed to allow mixed strategy equilibria, some of which are trembling-hand perfect (Selten 1975), while others are not (i.e., they are "imperfect equilibria"). Behavior in games with mixed strategy equilibria is often difficult to

[19] We searched using the four levels of reasoning from level-0 to level-3, because most studies find that these four levels are sufficient to explain the data (see Costa-Gomes and Crawford 2006 and Crawford 2013). This seems to be confirmed by the fact that highest level in the level-k mixtures that we identify is level-2. None of the identified mixtures contains level-3 types.

**Table 4**
Squared error minimizing level-k mixtures.

| | Level 0 | Level 1 | Level 2 | Level 3 |
|---|---|---|---|---|
| Short-run strangers treatment (all 50 periods) | 0.65 | 0.35 | 0.00 | 0.00 |
| Long-run strangers treatment (first 50 periods) | 0.65 | 0.30 | 0.05 | 0.00 |
| Long-run strangers treatment (last 50 periods) | 0.60 | 0.40 | 0.00 | 0.00 |
| Long-run partners treatment (first 50 periods) | 0.75 | 0.25 | 0.00 | 0.00 |
| Long-run partners treatment (last 50 periods) | 0.65 | 0.35 | 0.00 | 0.00 |

analyze, because observed choices do not generally reveal the underlying behavior strategies of the players. We used a setup with 100 plays per period, eliciting players' behavior strategies by allowing them to decide how many of the 100 plays they would choose *R*, with the remaining plays being *L* choices. In our short-run strangers treatment, subjects participated in 50 periods (of 100 plays each), where across periods they faced varying anonymous other players. In our long-run strangers treatment, we increased the number of periods to up to 250. In our long-run partners treatment, subjects also played up to 250 periods, but repeatedly interact with the same two other subjects (each with a fixed player type) throughout the session.

Our experimental implementation allows the play of all equilibrium strategy profiles. However, equilibrium profiles are rarely observed in the data and the relative frequency of the trembling-hand perfect equilibrium outcomes declines with repetition.[20] Although the "unreasonable" imperfect equilibrium outcomes (Selten 1975) are obtained more frequently than the trembling-hand perfect equilibrium outcomes, we also find only limited support for the imperfect equilibrium in our data.[21] The most frequently observed strategy profile in our short-run strangers treatment involves the play of the imperfect equilibrium strategies by players 1 and 3, but we rarely observe the corresponding equilibrium strategy of player 2. Since players 1 and 3 in these cases cannot observe the deviation of player 2′s strategy from equilibrium play, they do not promptly react with a best response to player 2′s non-equilibrium play. In the long-run strangers treatment, we most frequently observe player 1 mixing almost evenly between *L* and *R*, balancing the interests of the other two players and reaching a Pareto optimal outcome. In the long-run partners treatment, the most frequently observed outcome is the imperfect equilibrium outcome that also satisfies the condition of Pareto optimality.

Due to the inherent imperfect observability of the mixed strategy choices, learning in games with mixed strategy equilibria is usually rather demanding. Starting on the presumption that this also holds true for the sophisticated Game of Selten's Horse, we conjectured that impulse response dynamics may provide a suitable approximation for the observed behavioral dynamics in this game, as they have in other settings (e.g., Goerg, Neugebauer, Sadrieh 2016). Impulse response is a deterministic simulation of one-step best-response dynamics interrelated with learning direction theory (Selten and Stoecker 1986). In fact, we find that impulse response trajectories are well in line with the observed dynamics of play in our short-run strangers treatment sessions.

However, in the short-run, neither the simulated impulse response nor the observed behavioral trajectories converge on a rest point. In the long-run, impulse response trajectories always converge on the perfect equilibrium, but they require at least 160 periods to reach the rest point. To our surprise, the observed behavior in the long-run sessions with up to 250 periods does not follow that path. Instead, in the long-run strangers treatment, we find a tendency of player 1 to choose *R* and *L* with almost equal probabilities, hence, deviating both from the perfect and the imperfect equilibrium play. This out-of-equilibrium behavior enables player 1 in the Game of Selten's Horse to balance the diverging interests of the other players and to reach Pareto efficient outcomes, in which the own payoffs and the mutual payoffs of all players together are higher than in the perfect equilibrium.

In the long-run partners treatment, observed behavioral trajectories have a tendency to move towards the imperfect equilibrium which is also a Pareto efficient outcome of the game, albeit with a different distribution of payoffs than those generally observed in the long-run strangers treatment. We conjecture that in the long-run Pareto efficient outcomes of the game exert a strong attraction on behavior, both in anonymously repeated interactions (i.e., strangers treatment) and in repeated games (i.e., partners treatment). While the point of attraction differs, depending on the treatment, Pareto efficiency (or joint payoff maximization) obviously has a strong appeal for the subjects in long-run interactions.

Comparing the predictive power of the impulse response dynamics to other models of choice adaptation, we investigate other learning models involving reinforcement, the level-k and the quantal response approaches as alternatives. The level-k approach has the peculiarity in the Game of Selten's Horse that behavior on all levels of rationality lead to one of three possible strategy choices, choosing either only *R*, only *L*, or exactly 50% of each. Moving through the levels, we find that the resulting best responses to the next

[20] In independent research, Berninghaus, Güth and Li (2012) study a closely related 3-player (one-shot) game employing the strategy method. In their experimental design, subjects could not choose pure strategies, but always had minimum trembles. In their recent study, Sadanand and Tsakas (2023) also let subjects play a game similar to the Game of Selten's Horse (with a slightly altered payoff structure), but only allow their subjects to submit a single pure strategy choice for each play of the game, instead of letting them choose a mixed strategy as in our experiment. Despite the differences between the three games and between the three experimental setups, equilibrium profiles are infrequently observed in all studies. This seems to underline that there is an inherent instability in the strategic situation represented by the Game of Selten's Horse that robustly distracts behavior from the proposed equilibria.

[21] In the long-run partners treatment, we find 20 percent equilibrium profiles with about 16 percent being imperfect and 4 percent being perfect. In the strangers treatments, we find substantially less equilibrium profiles, with only about 3 percent in the long-run strangers treatment (almost all of them being imperfect) and about 4 percent in the short-run strangers treatment (a little more that 3 percent being imperfect and the rest being perfect).

lower show a complex cycle. Our simultaneous estimation results show all subjects at low levels of rationality without the dynamic behavioral adaptation found in some other games (Crawford, Costa-Gomes, and Iriberri 2013), which is possibly owed to the complexity of the game of Selten's Horse.

Similarly, the choice trajectories of quantal response approach (McKelvey and Palfrey 1995) that result from a decreasing level of noise in the model, only partially capture the main features of the observed trajectories in our experiment. The quantal response model predicts behavior to converge on the trembling-hand perfect equilibrium as noisy behavior diminishes over time. However, estimating the noise levels in a quantal response model, we find that noise levels decrease only early on in the interaction and only to a very small degree. Even in the long-run sessions, the estimated noise levels at the end of the experiment are about 16 times higher than the quantal response model would require for convergence.

We provide simulations of the reinforcement learning model (Erev and Roth 1998) and a mixed model combining reinforcement and impulse response learning (Camerer, 2003). Our findings indicate that, while the reinforcement and mixed learning trajectories are not closer than the impulse response trajectories to the observed behavior in the short-run and long-run strangers treatment, they do provide a better prediction than the impulse response trajectories in the long-run partners treatment. Both the impulse response simulation and the reinforcement learning model capture some important aspects of behavior in the Game of Selten's Horse, but neither can fully account for the dynamics of behavior.

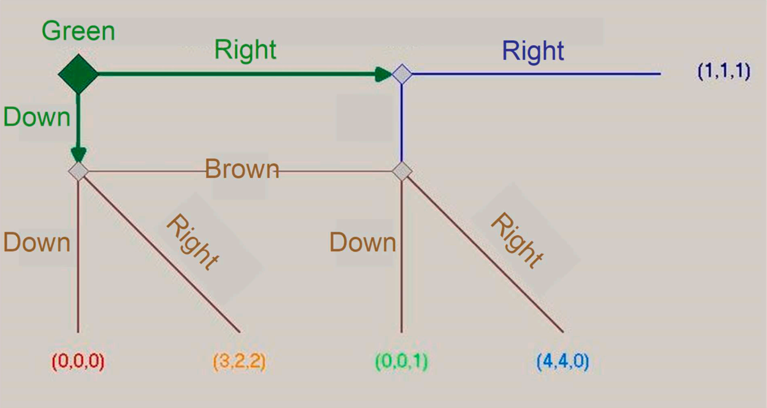
It seems that both impulse response and reinforcement learning – as well as most of the other behavioral choice adaptation models – only focus on specific driving forces, but miss out on others. Especially the way subjects learn to assess mixed strategies in the long-run seems to affect their choices in a more complex manner than the behavioral choice adaptation models currently consider. We conclude that more research is needed to enhance these models, enabling them to also capture the intricate driving forces in complex games with highly asymmetric mixed strategy equilibria. We find, the Game of Selten's Horse fits perfectly in this category of games, providing a challenging experimental paradigm for behavioral game theory.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

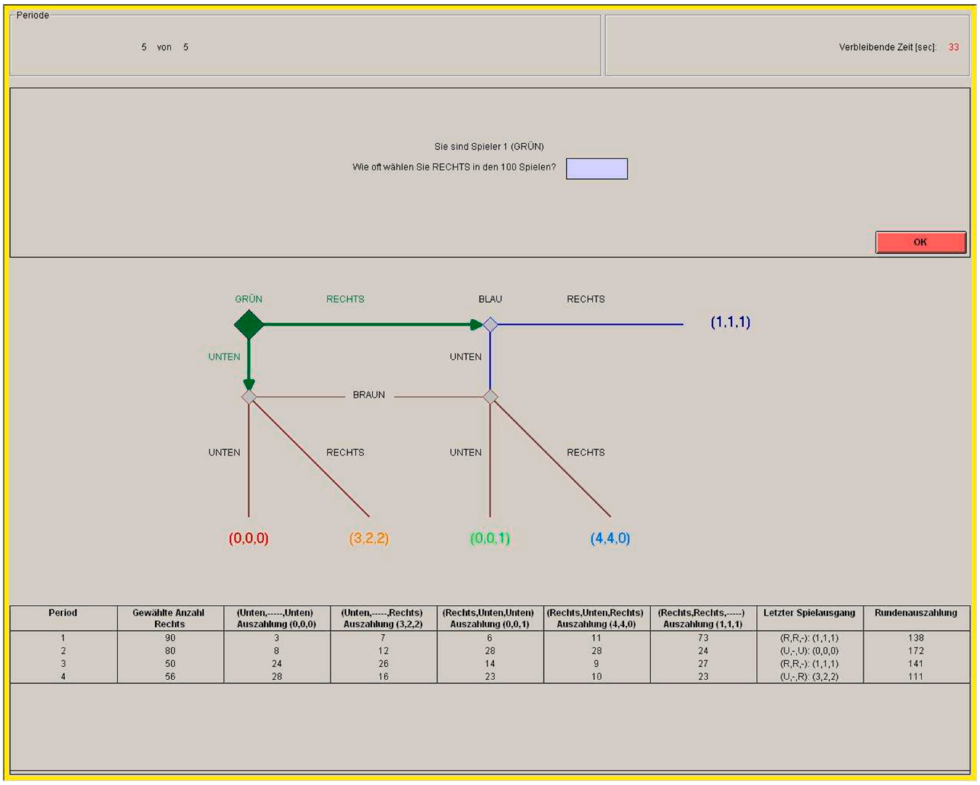
## Appendix 1. Instructions (short-run treatment)

- This experiment involves 50 rounds.
- In each round, you make decisions for 100 three-person games, which are played simultaneously.
- In every round you are matched and play with two other participants in the experiment. The matching in each round are random. The probability to play in consecutive rounds in the same group is small.
- The three players in each game have different positions. Players are either "Green", "Blue" or "Brown". At the beginning of the sequence each participant is assigned to one of three roles. You keep this role during the entire sequence.
- The picture shows the game being played 100 times in each round. The diamonds and lines refer to the decisions of the players. The numbers in parentheses indicate the payoffs of the players, the payments are ordered as follows (Green, Blue, Brown).



- Each player decides in each round, in how many of the 100 games of the round, he / she chooses "right" and in how many of the 100 games he / she chooses "Down". If the player, for example, chooses to play 80 games "Right", then s/he will play in 80 of the 100 games "right" and in 20 of 100 games "down". If he / she chooses to play in 20 games, "Right", then s/he will play in 20 of the 100 games "right" and in 80 of 100 games "down".
- At the end of each round you will be informed about the payoffs of all three participants in your group in the 100 games. (See the following figure.)
- [average pay condition:] You earn your average payoff in the 100 games of the round.
- [random pay condition:] You earn one of your payoffs resulting in a single game of the round. This game will be chosen randomly in each round.

- You will receive € 0.20 per point, and will be paid all your round payoffs.



Periode

5 von 5

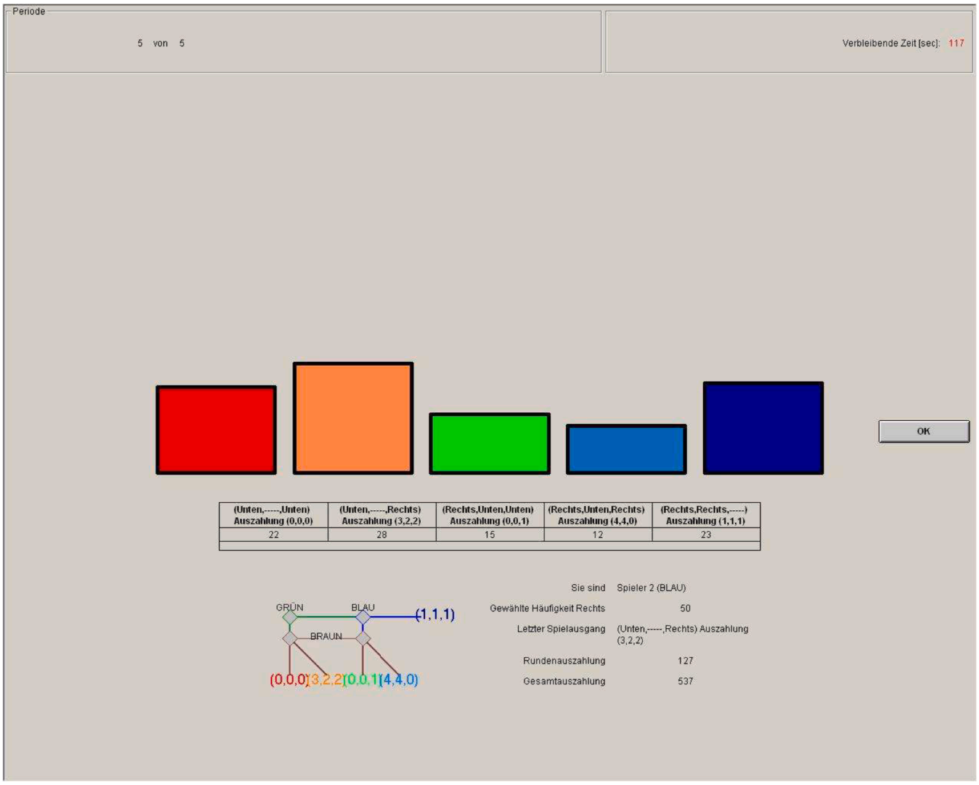
Verbleibende Zeit [sec]: 33

Sie sind Spieler 1 (GRÜN)

Wie oft wählen Sie RECHTS in den 100 Spielen?

OK

| Period | Gewählte Anzahl Rechts | (Unten,-----,Unten) Auszahlung (0,0,0) | (Unten,-----,Rechts) Auszahlung (3,2,2) | (Rechts,Unten,Unten) Auszahlung (0,0,1) | (Rechts,Unten,Rechts) Auszahlung (4,4,0) | (Rechts,Rechts,-----) Auszahlung (1,1,1) | Letzter Spielausgang | Rundenauszahlung |
|---|---|---|---|---|---|---|---|---|
| 1 | 90 | 3 | 7 | 6 | 11 | 73 | (R,R,-): (1,1,1) | 138 |
| 2 | 80 | 8 | 12 | 28 | 28 | 24 | (U,-,U): (0,0,0) | 172 |
| 3 | 50 | 24 | 26 | 14 | 9 | 27 | (R,R,-): (1,1,1) | 141 |
| 4 | 56 | 28 | 16 | 23 | 10 | 23 | (U,-,R): (3,2,2) | 111 |



Periode

5 von 5

Verbleibende Zeit [sec]: 117

OK

| (Unten,-----,Unten) Auszahlung (0,0,0) | (Unten,-----,Rechts) Auszahlung (3,2,2) | (Rechts,Unten,Unten) Auszahlung (0,0,1) | (Rechts,Unten,Rechts) Auszahlung (4,4,0) | (Rechts,Rechts,-----) Auszahlung (1,1,1) |
|---|---|---|---|---|
| 22 | 28 | 15 | 12 | 23 |

| | |
|---|---|
| Sie sind | Spieler 2 (BLAU) |
| Gewählte Häufigkeit Rechts | 50 |
| Letzter Spielausgang | (Unten,-----,Rechts) Auszahlung (3,2,2) |
| Rundenauszahlung | 127 |
| Gesamtauszahlung | 537 |

## Appendix 2. Tables

**Table A2.1**
Mean squared deviations of simulation versus choice trajectories.

| ind. obs. | **short-run strangers** | | | | **long-run strangers** | | | | **long-run partners** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | IR | RL | one-step RL | mixed model | IR | RL | one-step RL | mixed model | IR | RL | one-step RL | mixed model |
| 1 | **277** | 873 | 877 | 879 | **146** | 219 | 232 | 254 | 1935 | 1809 | 1748 | **1732** |
| 2 | 459 | **327** | 329 | 332 | 231 | **144** | 258 | 201 | 241 | **129** | 249 | 249 |
| 3 | **157** | 352 | 403 | 381 | 702 | 759 | **553** | 555 | 1961 | 974 | **884** | 1129 |
| 4 | **127** | 196 | 228 | 222 | 473 | **227** | 611 | 648 | 661 | **460** | 687 | 717 |
| 5 | **514** | 988 | 998 | 971 | 1414 | **543** | 1023 | 1020 | 358 | **38** | 48 | 49 |
| 6 | **160** | 196 | 217 | 213 | 410 | 680 | **352** | 326 | 405 | 248 | **50** | 50 |
| 7 | **933** | 1176 | 1287 | 1302 | 1002 | **284** | 356 | 365 | 651 | 906 | **449** | 449 |
| 8 | **64** | 100 | 106 | 100 | 365 | **306** | 501 | 487 | 796 | 541 | **313** | 324 |
| 9 | **1335** | 1568 | 1720 | 1705 | 1112 | **709** | 1049 | 1034 | 609 | **174** | 446 | 466 |
| 10 | 254 | 108 | **79** | 84 | 210 | 367 | 195 | **171** | 929 | **533** | 1027 | 817 |
| 11 | 375 | **129** | 150 | 146 | **192** | 441 | 294 | 279 | 787 | 481 | **428** | 454 |
| 12 | 722 | **585** | 607 | 603 | **505** | 505 | 923 | 917 | **1098** | 2281 | 2281 | 2281 |
| 13 | 939 | **862** | 868 | 864 | **250** | 619 | 440 | 372 | **375** | 468 | 589 | 598 |
| 14 | **151** | 232 | 265 | 242 | 198 | **145** | 249 | 230 | 3487 | **3127** | 3127 | 3127 |
| 15 | 668 | **630** | 661 | 659 | **343** | 467 | 361 | 360 | 780 | 505 | **258** | 281 |
| 16 | **314** | 373 | 381 | 377 | **69** | 270 | 126 | 136 | 303 | 47 | **31** | 31 |
| 17 | **349** | 711 | 723 | 671 | 1013 | **717** | 747 | 733 | 1075 | 975 | 503 | **462** |
| 18 | 1255 | 960 | **871** | 897 | 247 | **149** | 212 | 229 | 3828 | **3548** | 4605 | 4571 |
| # min | 11 | 5 | 2 | 0 | 6 | 9 | 2 | 1 | 2 | 7 | 7 | 2 |

Notes.
The mean squared deviations are average squared deviations of the simulation predictions from observed choices.
The points of estimation are the ten-period averages, each a number between 0 (L) and 100 (R).
For each player type, we compute the sum of squared deviations at each estimation points and report the average of these sums of squared deviations across all three player types.
There are 18 independent observations in each treatment, i.e., 54 independent observations in total.
The minimum of the mean squared deviations for each independent observation is in **bold face**.
# min = number of independent observations with the lowest squared deviations in this column.
IR = Impulse Response; RL = Reinforcement Learning.
mixed model = mixture of IR and RL corresponding to EWA (Camerer and Ho, 1999).

## Appendix 3. Figures

Fig. A31,Fig. A32,Fig. A33,Fig. A34,Fig. A35,Fig. A36,Table A21

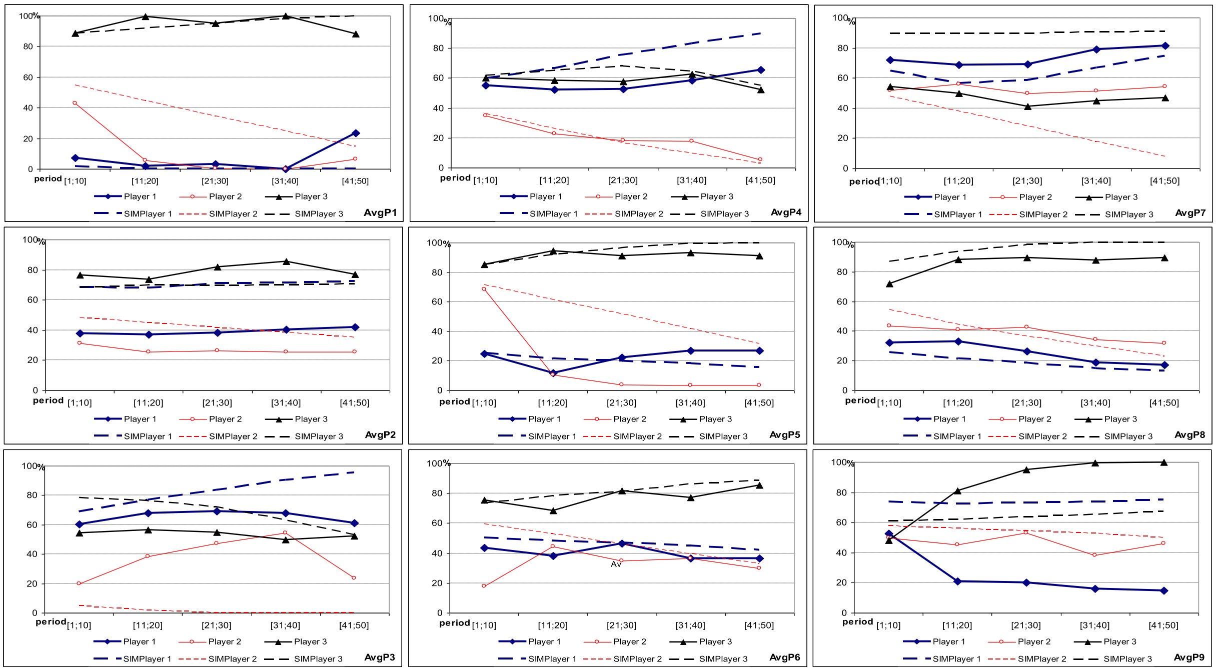


**Fig. A3.1.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay condition: average behavior strategies, probability of playing *R*, over 10 periods.

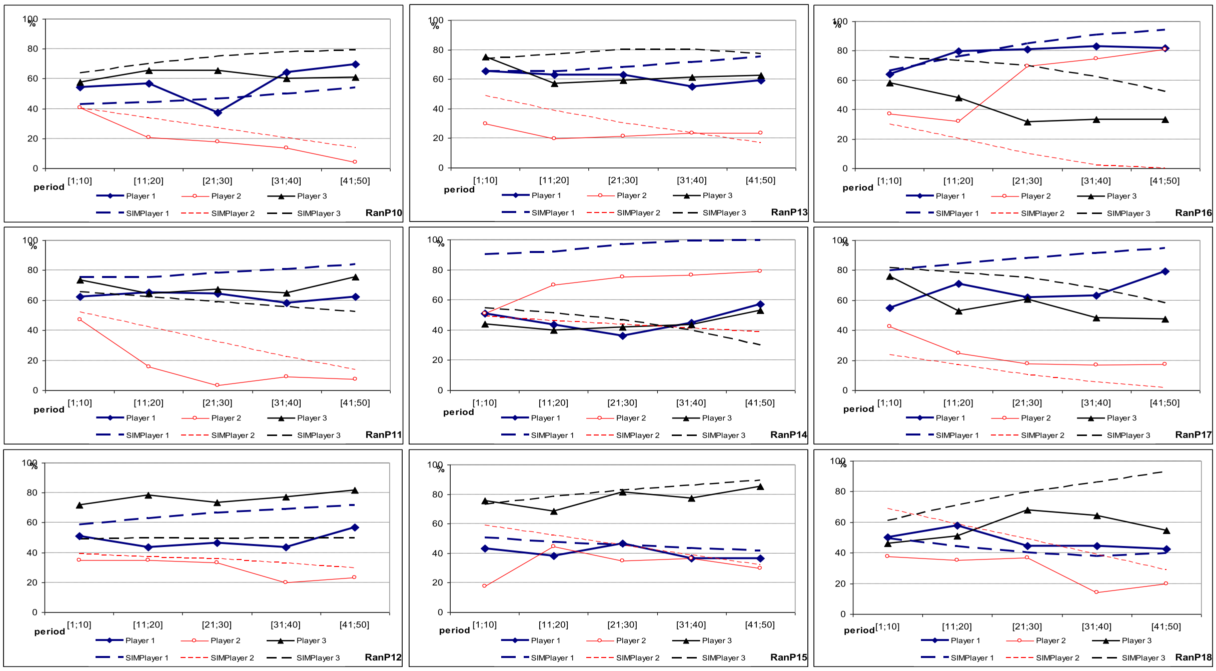


**Fig. A3.2.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay condition: average behavior strategies, probability of playing *R*, over 10 period.

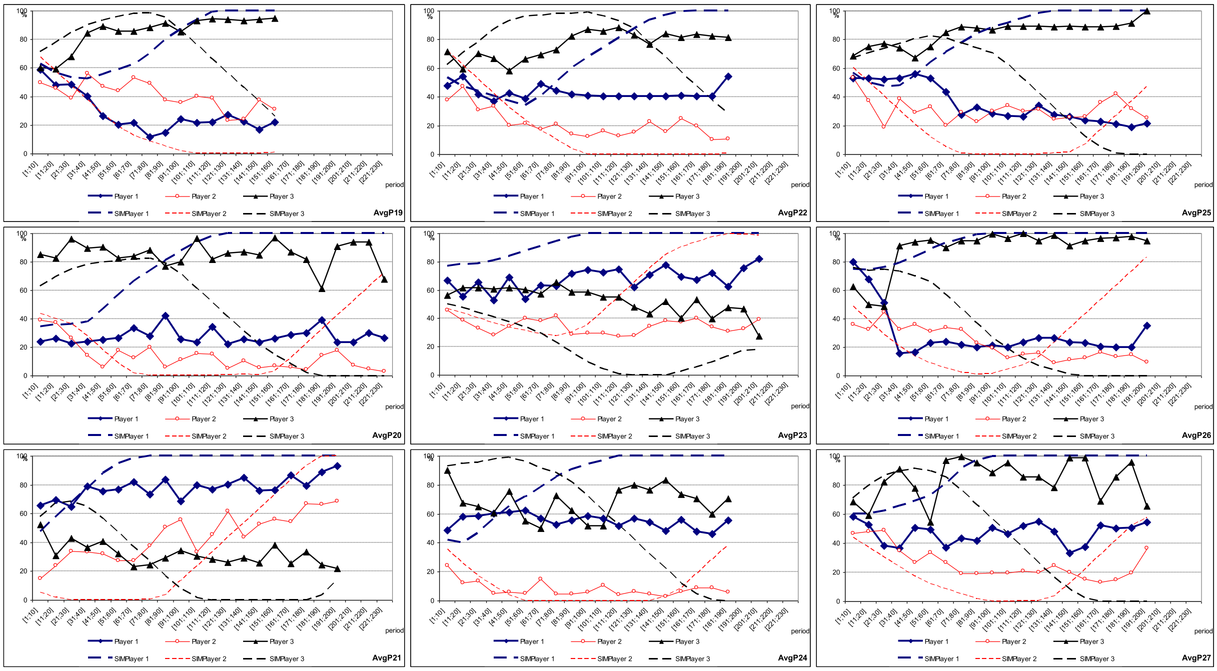


**Fig. A3.3.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay condition of the long-run strangers experiment: average behavior strategies, probability of playing *R* over 10 periods.

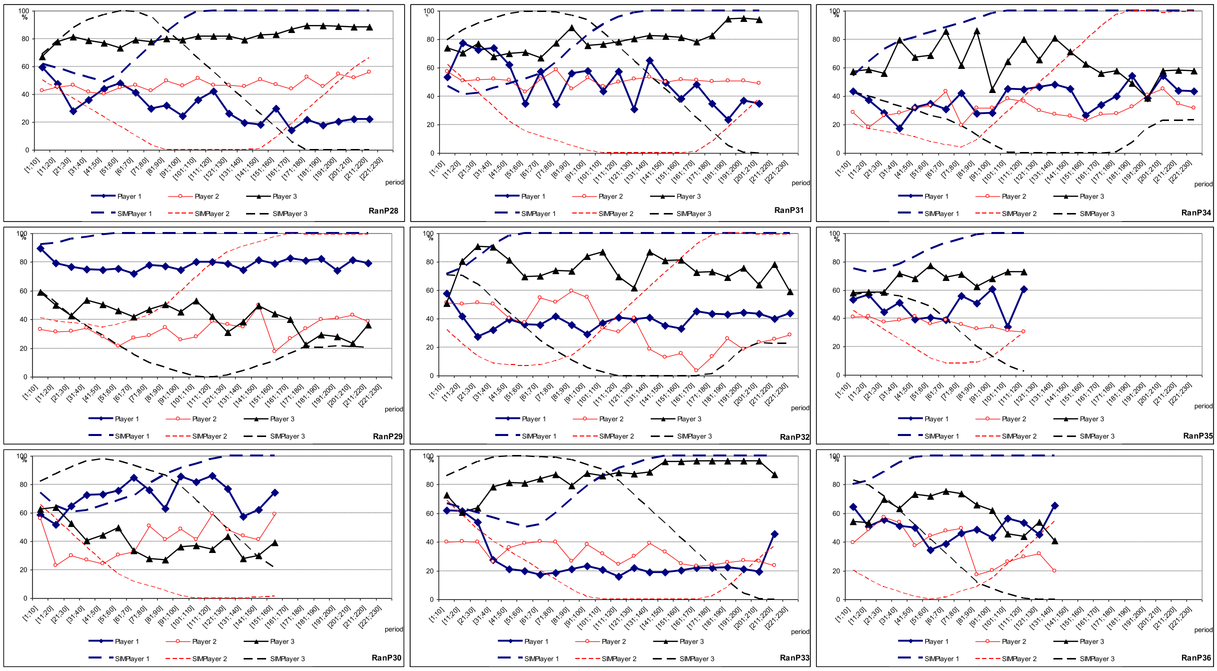


**Fig. A3.4.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay condition of the long-run strangers experiment: average behavior strategies, probability of playing *R* over 10 periods.

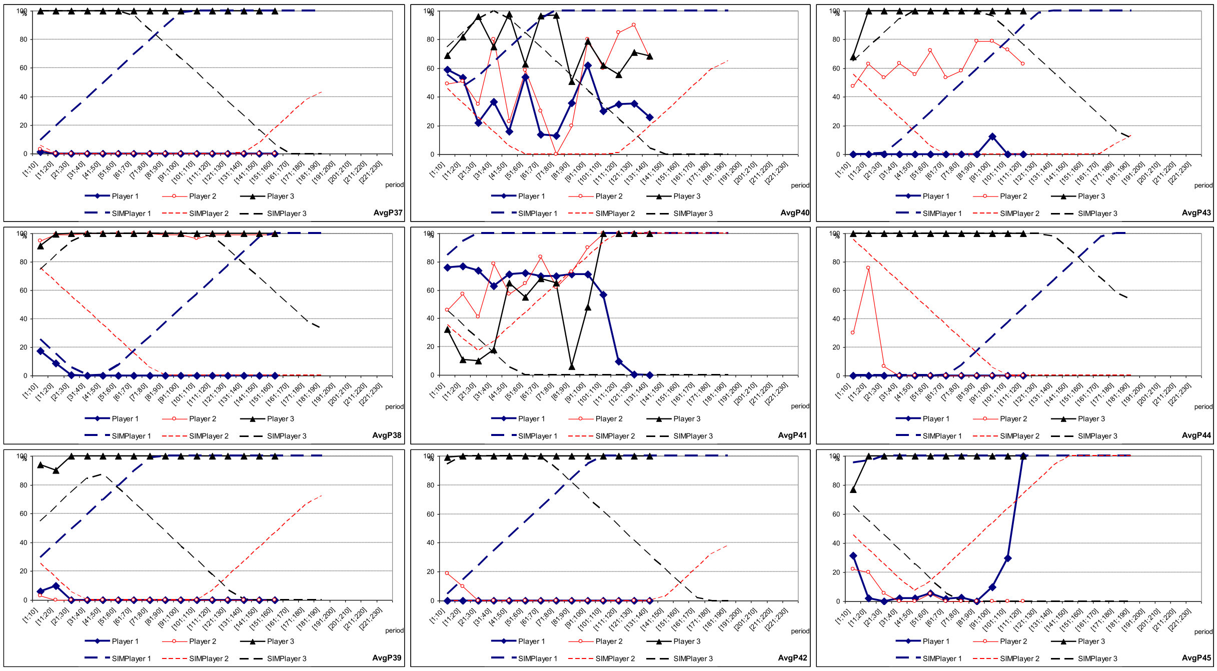


**Fig. A3.5.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay condition of the long-run partners experiment: average behavior strategies, probability of playing *R* over 10 periods.

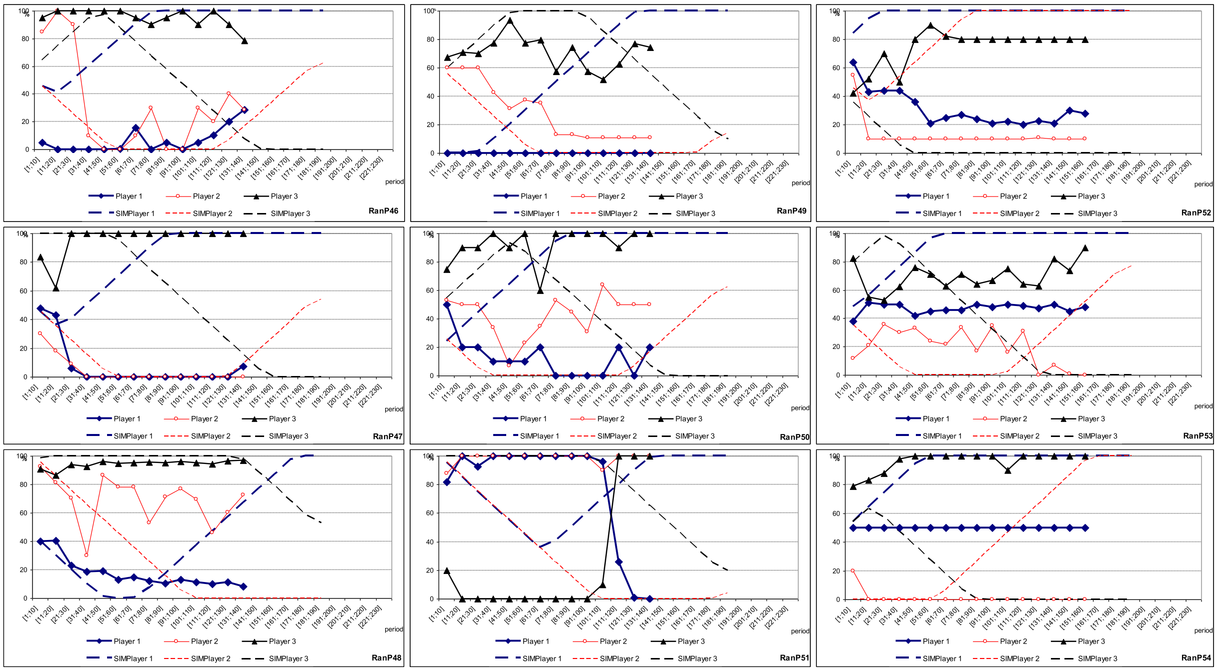


**Fig. A3.6.** Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay condition of the long-run partners experiment: average behavior strategies, probability of playing *R* over 10 periods.

## Data availability

Data will be made available on request.

## References

Binmore, K., 1987. Modeling rational players: part I. Econ. Philos 3 (2), 179–214.

Berninghaus, S., Güth, W., Li, K.K., 2012. Approximate Truth of perfectness: An experimental Test (41). Working Paper series in economics.

Camerer, C., 2003. Behavioral Game theory: Experiments in Strategic Interaction. Princeton University Press.

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

Capra, C.M., Goeree, J.K., Gomez, R., Holt, C.A., 1999. Anomalous behavior in a traveler's dilemma? Am. Econ. Rev 89 (3), 678–690.

Cheung, Y.-W., Friedman, D., 1997. Individual learning in normal form games: some laboratory results. Games Econ. Behav 19 (1), 46–76.

Chmura, T., Goerg, S.J., Selten, R., 2012. Learning in experimental 2× 2 games. Games Econ. Behav 76 (1), 44–73.

Chmura, T., Güth, W., 2011. The minority of three-game: an experimental and theoretical analysis. Games 2 (3), 333–354.

Costa-Gomes, M.A., Crawford, V.P., 2006. Cognition and behavior in two-person guessing games: an experimental study. Am. Econ. Rev 96 (5), 1737–1768.

Crawford, V.P., 2013. Boundedly rational versus optimization-based models. J. Econ. Lit 51 (2), 512–527.

Crawford, V.P., Costa-Gomes, M.A., Iriberri, N., 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. J. Econ. Lit 51 (1), 5–62.

Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev 88 (4), 848–881.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experim. Econ. 10 (2), 171–178.

Friedman, D., Oprea, R., 2012. A continuous dilemma. Am. Econ. Rev 102 (1), 337–363.

García-Pola, B., Iriberri, N., Kovářík, J., 2020. Non-equilibrium play in centipede games. Games Econ. Behav 120, 391–433.

Goeree, J.K., Holt, C.A., 1999. Stochastic game theory: for playing games, not just for doing theory. Proc. National Acad. Sci 96 (19), 10564–10567.

Goerg, S., Neugebauer, T., Sadrieh, A., 2016. Impulse response dynamics in weakest link games. German Econ. Rev 17 (3), 284–297.

Ho, T.H., Camerer, C.F., Chong, J.-K., 2007. Self-tuning experience weighted attraction learning in games. J. Econ. Theory 133 (1), 177–198.

McKelvey, R.D., Palfrey, T.R., 1995. Quantal response equilibria for normal form games. Games Econ. Behav 10 (1), 6–38.

McKelvey, R.D., McLennan, A.M., & Turocy, T.L. (2006). Gambit: software tools for game theory.

Nagel, R., 1995. Unraveling in guessing games: an experimental study. Am. Econ. Rev 85 (5), 1313–1326.

Neugebauer, T., Selten, R., 2006. Individual behavior of first-price auctions: the importance of information feedback in computerized experimental markets. Games Econ. Behav 54 (1), 183–204.

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

Ockenfels, A., Selten, R., 2005. Impulse balance theory and feedback in first-price auctions. Games Econ. Behav 51, 155–170.

Sadanand, A., Tsakas, N., 2023. Selten's Horse an Experiment on Sequential Rationality, 4427474. Available at SSRN.

Selten, R., 1965. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. Zeitschrift. für. die. gesamte. Staatswissenschaft, 121, 301–327 and 667-689.

Selten, R., 1973. A simple model of imperfect competition, where 4 are few and 6 are many. Int. J. Game Theory 2, 141–201.

Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. Int. J. Game Theory 4 (1), 25–55.

Selten, R., 2004. Learning direction theory and impulse balance equilibrium. In: Friedman, D., Cassar, A. (Eds.), Economics Lab—An Intensive Course in Experimental Economics. Routledge, NY, pp. 133–140.

Selten, R., Abbink, K., Cox, R., 2005. Learning direction theory and the winner's curse. Exper. Econ 8 (1), 5–20.

Selten, R., Buchta, J., 1999. Experimental sealed bid first price auctions with directly observed bid functions. In: Rapoport, A., Budescu, D.V., Erev, I., Zwick, R. (Eds.), Games and Human Behavior: Essays in Honor of Amnon Rapoport, pp. 101–116.

Selten, R., Chmura, T., 2008. Stationary Concepts for Experimental 2x2-Games. Am. Econ. Rev 98 (3), 938–966.

Selten, R., Chmura, T., Goerg, S.J., 2011. Stationary concepts for experimental 2x2-games: reply. Am. Econ. Rev 101 (2), 1041–1044.

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoners' dilemma supergames: a learning theory approach. J. Econ. Behav. Organ 7, 47–70.

Stahl, D.O., Wilson, P.W., 1995. On players' models of other players: theory and experimental evidence. Games. Econ. Behav 10 (1), 218–254.