Taming Selten's Horse with Impulse Response[‡]

Tibor Neugebauer*, Abdolkarim Sadrieh**, Reinhard Selten***

* University of Luxembourg
** University of Magdeburg
*** University of Bonn

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 indicates lower error rates in later than in earlier periods thus also suggesting the trembling-hand perfect equilibrium for the long run. In the long run (more than 200 periods), however, behavior seems to settle at a non-equilibrium distribution of strategies that supports efficient outcomes or low level-k of strategic sophistication instead of converging to the trembling-hand perfect equilibrium.

***JEL code*:** 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

[‡] 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 (F2R-368 LSF-PMA-13SYSB).

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 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 mechanisms for situations with a multiplicity of equilibrium points. However, their empirical relevance has yet to be shown.

In our study, we have conducted experiments of the Game of Selten's Horse to check in how far empirical evidence can support the trembling-hand perfect equilibrium. To our great 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 much better the observed pattern of play 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, 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 our first experiment, closer than the reinforcement learning trajectories do. Thus, we tentatively conclude that the perfect equilibrium dynamics are at work, but that full convergence to the set of perfect equilibrium points may take more repetitions. In a second experiment, we extend the number of repetitions to more than 200 periods, but do not find convergence to the perfect equilibrium, as we had expected. Instead, we find that behavior seems to settle at a non-equilibrium distribution of strategies that shows a low level of strategic sophistication in the level-k model, but is supported by high levels of total payoffs leaning towards an overall efficient play. When subjects are partners in the repeated game, particularly, high levels of payoffs are prevalent.

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 equilibrium. Section 3

describes our experimental design, and section 4 reports the static test of the perfect equilibrium points. Section 5 discusses learning direction theory, impulse balance, impulse response as well as reinforcement learning trajectories, and compares the simulated trajectories to the data. Section 6 offers insights on alternative models and explanations, especially the quantal response (McKelvey and Palfrey 1995) and the level-k (Crawford 2013) models. Section 7 provides a robustness check of the results in long-term settings, before section 8 concludes the paper.

2 Theoretical considerations

Selten's Horse is depicted in Figure 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.

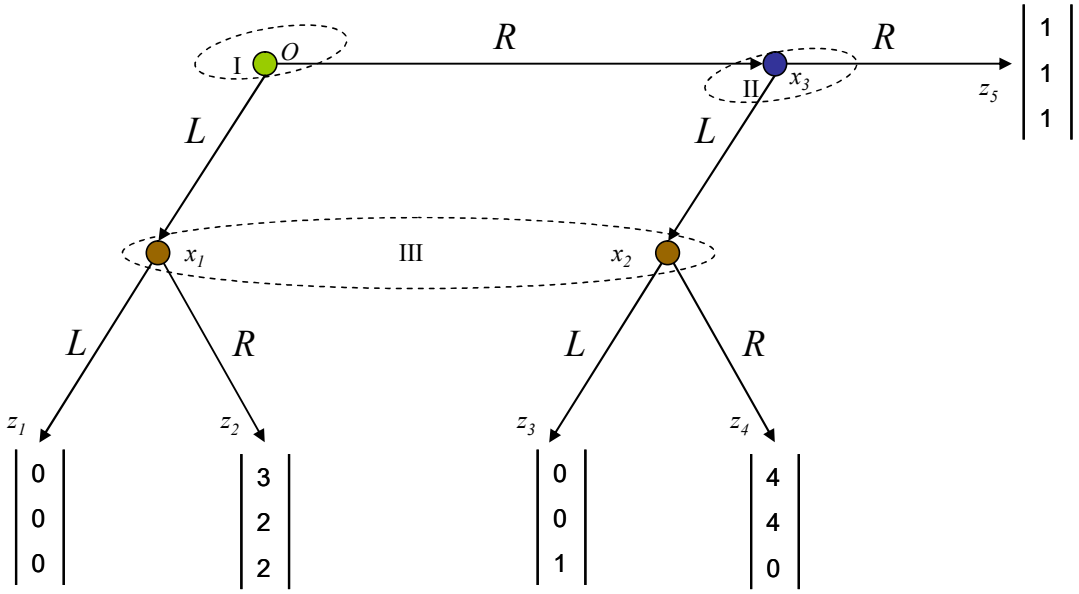


Figure 1. Selten's Horse

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 in the Game of Selten's Horse are thus described as follows:

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

$$p_2 = \begin{cases} 1, & < \\ [0,1] & \text{if } p_3 = 0.25 \\ 0 & > \end{cases}$$

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

There are two types of equilibrium points.

Trembling-hand perfect equilibrium points: $p_1 = 1, p_2 = 1, 0 \le p_3 \le \frac{1}{4}$ (1)

Imperfect equilibrium points: $p_1 = 0, \frac{1}{3} \le p_2 \le 1, p_3 = 1$ (2)

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 equilibrium type, followed by the discussion of the trembling-hand perfect equilibrium points.

Imperfect equilibrium points 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 his 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 his information set (node x_3 in Figure 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 he believes that player 3 will choose R according to the equilibrium point, then his best response is L where he will receive a payoff of 4 instead of R with a payoff of 1. The same reasoning also applies to the other mixed strategy imperfect equilibrium points.

This is different in the trembling-hand perfect equilibrium points, where the information set of player 3 should not be reached. Even if players 1 and 2 make mistakes so that player 3's information set is reached, the trembling-hand perfect equilibrium strategies still maximize player 3's expected payoff.

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 equilibrium points are perfect equilibria, i.e. limit points of test sequences of the perturbed game.[1]

3 Experimental design

To test the prediction of trembling-hand perfect equilibrium theory, we conducted six computerized (Fischbacher 2007) experimental sessions with Selten's Horse at the MaxLab of the University of Magdeburg in 2010. Each session involved 27 subjects, split in 3 independent groups of 9. A group of 9 consisted of 3 subjects of each player type. Subjects maintained their player type throughout the session. Groups of 9 subjects interacted over 50 *periods* of random matching. Per period, 3 subgroups were randomly matched in each group. Subjects were not told the group size, but they were informed that the likelihood of being matched with the same two subjects in consecutive periods was small.

Each period contained 100 random *plays* of the Game of Selten's Horse in each sub-group. In every period, each subject of player type *i* chose the *relative frequency* f_i of playing *R* in the 100 plays, knowing that with the remaining frequency, $(1 - f_i)$, *L* would be played. This choice represents our experimental implementation of the behavior strategies p_i in the Game of Selten's Horse. A series of actions of *R* and *L* in the 100 plays was randomly drawn

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

(without replacement) for each player *i* according to f_i. The 100 play outcomes were determined by combining the action series of the three players in the subgroup.

There are two standard implementations of mixed-strategy payoffs in game theory experiments, which we apply as treatment variations. The players either receive the average payoff of all plays or the payoff of one randomly selected play.[2] We vary the payment modalities in our treatments, accordingly. In our *Average Pay Treatment,* the payment in a period is equal to the average payoff over the 100 plays. In the *Random Pay Treatment,* the period payment is equal to the outcome realized in one of the 100 plays in a 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 both treatments. The outcomes in the 100 plays of a period were presented on the screen in a histogram that showed the observed frequencies of each possible outcome of the game $z_1,...,z_5$ in the 100 plays. The subjects also learned in both treatments the outcome of one particular play. Subjects additionally received a record of past period earnings and total earnings.

4 General observations

The experiment involves 18 independent observations, 9 in Average Pay and 9 in Random Pay. Subjects interacted in 3×3 groups in 50 consecutive periods of 100 plays each. In total, 162 subjects participated in the experiment submitting a total of 3×2,700 behavior strategies. By participating in the experiment, a subject achieved an average payoff of € 16.10. Subjects received no show-up fee. Experimental sessions were completed within an hour, including the reading of the instructions.

Observation 1: The data show no significant treatment effect.

Figure 2 plots the behavior strategies submitted by players 1 to 3 over the course of the experiment. As one can easily see from the chart, the differences between the treatments are small and the average behavior strategies are quite stable over time in both treatments. We find no significant differences in decisions or outcomes between treatments (Mann-Whitney U-Test, $\alpha = 0.1$, two-tailed).

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

The following behavior strategies represent the average choice per treatment;

$f_1^{AvgP} = .425, f_2^{AvgP} = .242, f_3^{AvgP} = .719$ in the Average Pay treatment and

$f_1^{RanP} = .534, f_2^{RanP} = .410, f_3^{RanP} = .637$ in the Random Pay treatment.

Figure 3 shows the observed outcome distributions of the two treatments, corresponding to the notation $\{z_1, z_2, \ldots, z_5\}$ in Figure 1. We find small, but insignificant differences between outcome frequencies across treatments. Even the largest treatment differences – as seen for the outcomes z_2 (0, ·, 1) and z_5 (1, 1, ·) – are not significant.

The observed average outcomes in Figure 3 correspond to observed average earnings of 2.26, 1.71, and 1.10 per period in Average Pay and of 1.88, 1.82, and .91 in Random Pay for players of type 1, 2 and 3, respectively. Comparing these results to the trembling-hand perfect equilibrium points, we find that the average earnings of player 3 are close to the equilibrium prediction. Players 1 and 2, however, earn in excess of this equilibrium prediction. All three players earn substantially less than in the imperfect equilibrium. Hence, even though the game structure provides incentives to select the trembling-hand perfect equilibrium, out of equilibrium play empirically seem to entail very few negative payoff effects.

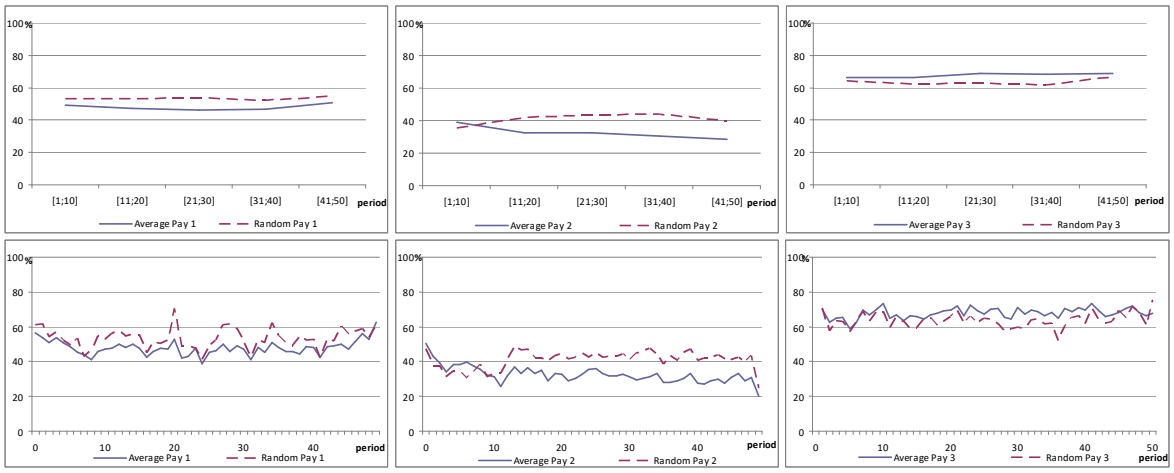


Figure 2 Average behavior strategies chosen by players 1 to 3 (left to right).

Solid line: Average Pay. Dashed line: Random Pay. Top: 10-period averages. Bottom: single-period average.

Table 1 records the cumulative probabilities of playing R for each player type,[3] and thus provides an overview of the observed individual strategies. The numbers show that the

[3] Overall periods, rounds, and players, the relative frequency of pure strategies was 46%, with 23% $p_i = 0$ and 23% $p_i = 1$. The remaining 54% were mixed strategies.

most frequent choices were pure strategies. Player 1 type subjects chose imperfect equilibrium strategy $p_1 = 0$ and the perfect equilibrium $p_1 = 1$ about equally frequently (about 20% each); player 2 type subjects most frequent choice was non-equilibrium strategy $p_2 = 0$ (38% of time); and player 3 type subjects chose to play the imperfect equilibrium strategy $p_3 = 1$ more frequently (about 30% of time) than any other strategy. Table 1 also indicates small differences in behavior between treatments. For instance, player 1 type subjects chose the pure strategy $p_1 = 1$ more frequently in Random Pay than in Average Pay, implying the difference in average outcomes z_2 and z_5 that can be seen in Figure 3.[4] Generally, we find no great differences in behavior between treatments which is overall good news for game theory. It suggests that both implementations, Average Pay and Random Pay, lead to similar results when using mixed strategies in the laboratory. From the recorded numbers in Table 1 the following observations are straightforward.

Observation 2: The trembling-hand perfect equilibrium strategy of player 3, $p_3 \leq \frac{1}{4}$, and the pure trembling-hand perfect equilibrium strategy of player 2, $p_2 = 1$, are observed infrequently.

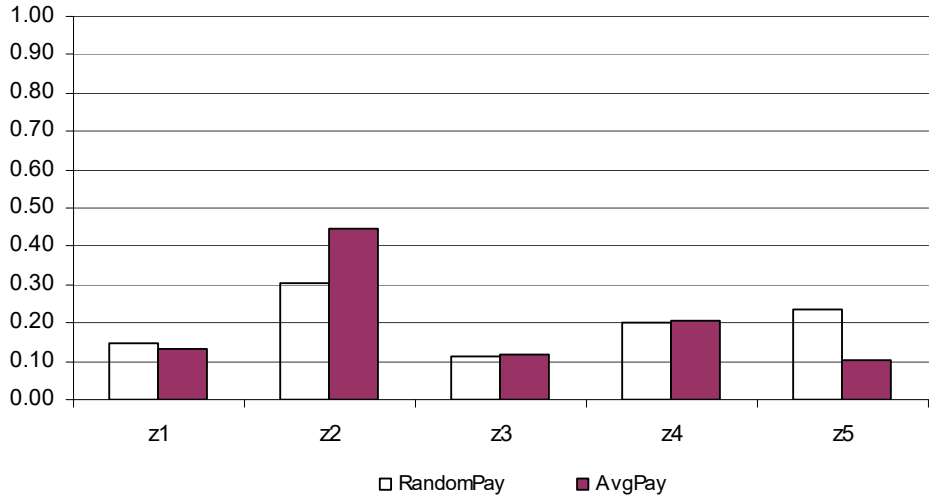


Figure 3 Average outcomes of play

[4] Player 3 subjects play more frequently left in Random Pay than in Average Pay, probably because they fear to regret their decision $p_3 > 0$, if z_4 is randomly chosen as final outcome. Similarly, player 2 and player 1 receive nothing if z_1 or z_3 is chosen as final outcome in Random Pay. The obvious response to a decrease in p_3 is thus an increase in p_1 and p_2.

Table 1 (see also Figure A1.1 in the appendix) shows that trembling-hand perfect equilibrium strategy is rarely played by player 3: less than 13% under Average Pay and less than 15% in the Random Pay treatment. The highest frequency of trembling-hand perfect equilibrium play by player type 3 is 24% in one independent group and the minimum is 1% in another. These observed frequencies are even below the frequencies expected by chance (26%). Hence, the reported averages indicate no support for the trembling-hand perfect equilibrium strategy of player 3.

According to the trembling-hand perfect equilibrium strategy, player 2 is expected to play *R* for sure, i.e., $p_2 = 1$. In contrast to this prediction, we only observe a corresponding action of player 2 in 11% of the choices. Hence, player 2 plays $p_2 = 1$ substantially less often than expected in the trembling-hand perfect equilibrium (100%), but substantially more often than suggested by a random choice from the entire action set, which sets $p_2 = 1$ as only one of the 101 possible levels {0.00, 0.01, 0.02, …, 0.99,1.00}.

Table 1. Cumulative distribution of subjects' behavior strategy

		p=0	[0,.25]	[0,.33]	[0,.50]	[0,.75]	[0,.99]	[0,1]
player 1	AvgPay	**0.271**	0.384	0.449	0.607	0.790	0.868	1
	RandomPay	**0.247**	0.312	0.330	0.470	0.629	0.801	1
	Overall	**0.259**	0.348**	0.389	0.539	0.710	0.834	1*
player 2	AvgPay	**0.443**	0.671	0.719	0.836	0.878	0.936	1
	RandomPay	**0.347**	0.497	0.530	0.636	0.690	0.809	1
	Overall	**0.395****	0.584**	0.624	0.736	0.784	0.873	1
player 3	AvgPay	0.061	0.128	0.147	0.270	0.443	0.616	**1**
	RandomPay	0.063	0.169	0.207	0.353	0.564	**0.806**	1
	Overall	0.062***	0.149	0.177	0.312	0.503	0.711	**1**

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

Observation 3: Player 3's imperfect equilibrium strategy $p_3 = 1$ is more frequently observed than the perfect equilibrium strategy $p_3 \leq \frac{1}{4}$.

Player 3 submitted $p_3 = 1$ in 38% on average of decisions under Average Pay and 23% under Random Pay, overall 30% of all decisions. This behavior is in line with the imperfect equilibrium points. The difference to the frequency of observed trembling-hand perfect equilibrium choices, $p_3 \leq \frac{1}{4}$, which are observed in 13% and 19%, overall 16%, of player 3 decisions, is significant at the 10% level.[5] Hence, we state the following.

Observation 4: The trembling-hand perfect equilibrium strategy profiles are less frequently observed in our data than the imperfect equilibrium strategy profiles.[6]

A related result is that the relative frequencies of outcomes z_2 and z_5 are also significantly different from one another (see figure 3); the p-value of the two-tailed Wilcoxon signed ranks test is .010. The imperfect equilibrium outcome is significantly more frequent in the data than the perfect equilibrium outcome.

In line with Selten's (1975) discussion of implausible behavior, we observe that player 2 chooses the imperfect equilibrium strategy, i.e. $\frac{1}{4} \leq p_2 \leq 1$, in the minority of the cases. Instead, player 2 chooses $p_2 = 0$ most of the time, particularly when the other two players play according to the imperfect equilibrium strategy. We observe this particular type of strategy profile (0,0,1) in about one quarter of the data. For reasons of completeness we also note that in 5% of the cases players 1 and 2 play in line with their trembling-hand perfect equilibrium strategy, but player 3 deviates by choosing $p_3 = 1$ instead of $p_3 \leq \frac{1}{4}$.

5 Best response dynamics

A closer look at the data suggests that subjects adjust their strategies in a best response manner. When player 1 increases his probability of playing R, player 3 frequently decreases the probability of choosing R and vice versa. When player 3 increases his probability of playing R, player 2 frequently decreases his probability of choosing R and vice versa. In the following we investigate the best-response dynamics more closely.

[5] The p-value of the two-tailed, one-sample Wilcoxon signed ranks test on the 18 independent observations is 0.089.

[6] Furthermore, according to the imperfect equilibrium strategy of player 2, the probability of playing *R* should be at least one third. In contrast, the data show that 61% of player 2's choices involved smaller probabilities than predicted in any of the imperfect equilibrium strategies. Again, by random choice from the entire action set we would expect a higher frequency of choices that are in line with the imperfect equilibrium strategy.

5.1 Learning direction theory

We apply learning direction theory (Selten and Stoecker 1986, Selten and Buchta 1999) to the data. According to learning direction theory, a player adjusts his behavior in hindsight in the direction of the ex-post best response or leaves it unchanged. Selten and Buchta (1999) illustrate learning direction theory by their analogy of an (autodidactic) marksman who learns how to hit a trunk with an 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 & Buchta 1999).

The ex-post best response function takes the other players' actions as given, and determines the best response to these given actions. We can use the best response function given in section 2. In each best response function, we replace the strategy choices of the other players by the last period's observed frequency of choosing *R*:

$$p_1 = \begin{cases} 1, & > \\ [0,1] & \text{if } 4(1-f_2^{t-1})f_3^{t-1} + f_3^{t-1} = 3f_3^{t-1} \\ 0 & < \end{cases}$$

$$p_2 = \begin{cases} 1, & < \\ [0,1] & \text{if } f_3^{t-1} = 0.25 \\ 0 & > \end{cases}$$

$$p_3 = \begin{cases} 1, & > \\ [0,1] & \text{if } 2(1-f_1^{t-1}) = f_1^{t-1}(1-f_2^{t-1}) \\ 0 & < \end{cases} \qquad (3)$$

The feedback information on the distribution of outcomes allows subjects to approximately infer the strategies of the others. According to learning direction theory, subjects are more likely to adapt their strategy in the direction of their ex-post best response than in another direction. This impulse to adapt the strategy rests when the subject has played the best response. In this case, direction learning theory predicts no change.

Observation 6: Subjects' behavior is in line with learning direction theory.

We can test learning direction theory on the individual or the group level. On individual level, we measure whether subject exhibit more changes in the predicted than in the opposite direction. Column 1 in Table 2 reports the number of subjects in each independent matching group, who make more changes in the predicted direction. In 15 of 18 independent matching groups (83%) the majority of subjects behave in line with learning direction theory. The players deviating from learning direction theory are almost equally spread over the player types. (See the footnote in Table 2.)

Table 2. Evidence in favor of learning direction theory

Group	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	Excess number of changes as predicted over changes in opposite direction	Excess number of responses in line rather than violating the prediction
AvgP1	9	9	24	41
AvgP2	9	9	44	188
AvgP3	8	9	63	172
AvgP4	7	8	7	90
AvgP5	2	3	-105	-84
AvgP6	8	8	126	140
AvgP7	8	8	47	105
AvgP8	7	8	26	118
AvgP9	5	8	3	171
RanP1	7	8	83	102
RanP2	3	5	-26	-2
RanP3	5	9	5	135
RanP4	7	9	15	133
RanP5	8	8	22	120
RanP6	4	8	24	177
RanP7	9	9	38	189
RanP8	7	7	38	58
RanP9	6	8	49	76
Total	119[a] (73%)	141[b] (87%)	483 (56%)	1929 (68%)

[a] 16, 14, 13 subjects of player type 1, 2, and 3 are at odds with learning direction theory.

[b] 7, 7, 7 subjects of player type 1, 2, and 3 are at odds with learning direction theory.

Instead of only counting the cases in which a subject makes a predicted or unpredicted change, we can also count the cases in which the subjects repeat their last choice. These cases are – strictly speaking – also in line with learning direction theory, which predicts either no change or a change in the direction of the best response. Column 2 in Table 2 reports the number of subjects in each matching group who exhibit a behavior that is in line with learning direction theory in this strict sense. Given these numbers, 17 of 18 independent matching groups (94%) have a majority of subjects making choices in line with the learning direction

theory. Moreover 15 of 18 independent matching groups (83%) score an 8 or a 9, i.e. have almost all nine members in line with learning direction theory. Again, we find no bias in the player types deviating from the learning direction theory.

On the group level, we can test for the excess of the number of changes in the predicted over the opposite direction. Alternatively, we also count in the cases without a change of the strategy as being in line with learning direction theory and examine the excess of the choices in line over those violating the learning direction theory. The columns 3 and 4 of Table 2 show the two scores for each of the independent matching groups. In total, 16 of 18 independent matching groups (89%) involve more changes (column 3) or more choices (column 4) in line with learning direction theory than in the opposite direction. Only two observations involve more changes or choices in the opposite direction than predicted by learning direction theory.[7] Overall, our data clearly support learning direction theory.[8]

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.[9] 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).[10]

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

[7] The probability of observing 2 or less failures in 18 observations is 0.0013 if both failure and success are equally likely. The Wilcoxon signed ranks test is also significant at the 1 percent level.

[8] Comparing the data to studies of less complex games, however, we find lowever agreement with learning direction theory. Neugebauer and Selten (2006), for example, report that only 8% of their subjects' behavioral patterns were at odds with learning direction theory.

[9] In most specifications, the impulses based on losses are weighted more strongly than those based on 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.

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

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* $\tilde{p}_{it}$ by one step. If no impulse is given, e.g. in the impulse balance point, the adjustment process rests.

$$\tilde{p}_{it} = \tilde{p}_{it-1} + r_i(t-1), \quad \textit{where} \tag{4}$$

$$r_i(t) = \begin{cases} .01, & \textit{if } p_{it} = 1 \wedge \tilde{p}_{it} < 1 \\ -.01, & \textit{if } p_{it} = 0 \wedge \tilde{p}_{it} > 0 \\ 0 & \textit{otherwise} \end{cases}$$

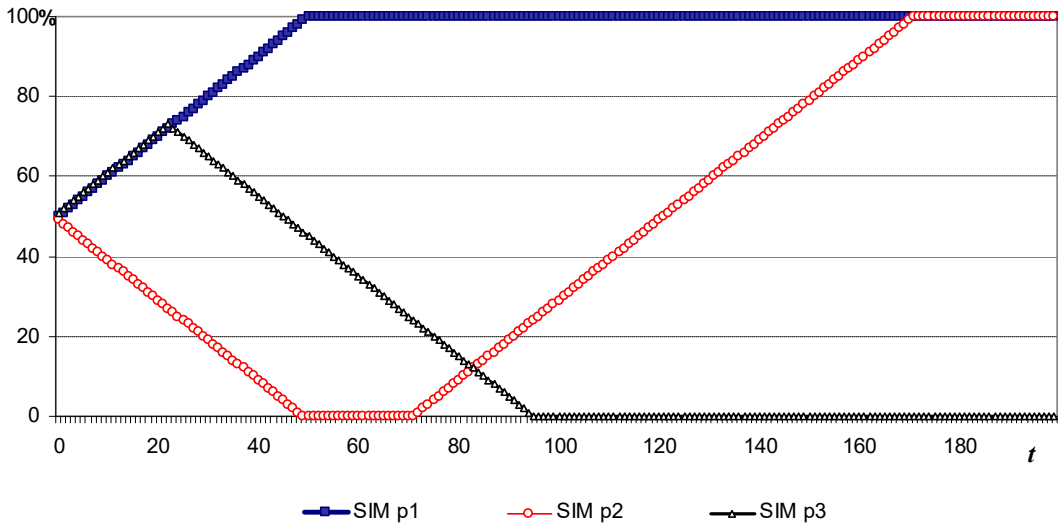


Figure 5 Impulse response trajectory of strategies chosen by players of type 1 (filled square), 2 (empty diamond), 3 (empty triangle) with initial profile (.5, .5, .5)

For given an initial strategy profile (.5, .5, .5), figure 5 exhibits the 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 perfect pure strategy equilibrium in the Game of Selten's Horse. This finding is curious, as the data show not much support of the perfect equilibrium and we have just shown that direction learning theory, which governs the dynamics of the impulse balance, is supported by the data.

We have a closer look at the dynamics within the data by studying the impulse response trajectories separately for each single 9-subject group. Our simulation ignores the actually received feedback of subjects or the original matchings after period 1, but we start

out the simulation at the initially observed strategy profiles and maintain the matching procedure within the 9-subject groups. In figures 6.1 (Average Pay) and 6.2 (Random Pay) we present the resulting simulation outcomes in a chart for each group jointly with the (smoothed) actually observed trajectories over the 50 repetitions. The observed behavior strategies of each player type are averaged over ten periods for each session, that is, each dot in the chart represents the average over 3×10 decisions of experimental subjects. For many groups the fit of the simulated trajectories is impressively close to the observed trajectories.

Observation 7: The impulse response trajectories are closer to the observed trajectories than chance.

As a benchmark, we conduct a Monte Carlo simulation with a one step adjustment by period, but where innovations are random. We compute the mean squared error on the described averages of 10 periods for each Monte Carlo simulation, and also for the impulse response simulation in 1,000 simulations. We find that for 14 sessions of 18 sessions the average mean squared error of the impulse response trajectory is smaller than the average mean squared error of the Monte Carlo simulation (see Table A1 in the appendix). The probability that 14 out of 18 or more successes would be drawn by chance is as low as .004. So we conclude that the fit of the observations by the impulse response trajectories is significantly better than chance.

Despite the fact that we do not show the long-term dynamics in the charts of figure 6, note that all our simulations converge on the impulse balance point. There are important differences in the number of required repetitions for the convergence to complete, depending on the initial strategy profile of the group. A comparison of figures 5 and 6 suggests that the observed trajectories need more time than the simulated trajectories. By the end of most sessions, the trajectory of player 3's strategy is still moving to R and the strategy of player 2 to L, corresponding to the very first part in Figure 5.

Observation 8: Reinforcement trajectories are closer to the observed trajectories than chance, but impulse response trajectories are even closer.

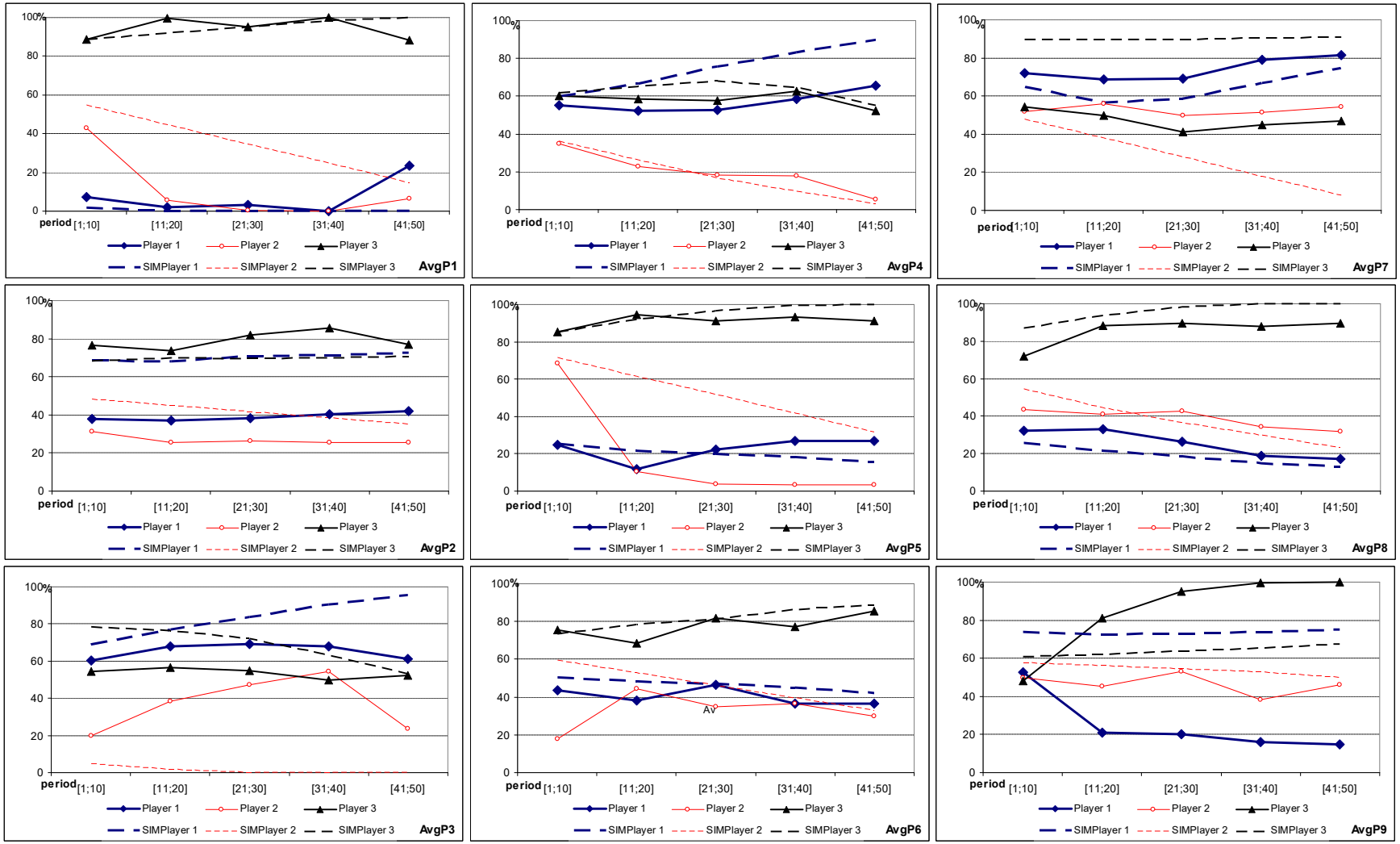


Figure 6.1 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay treatment: average behavior strategies, probability of playing R, over 10 periods

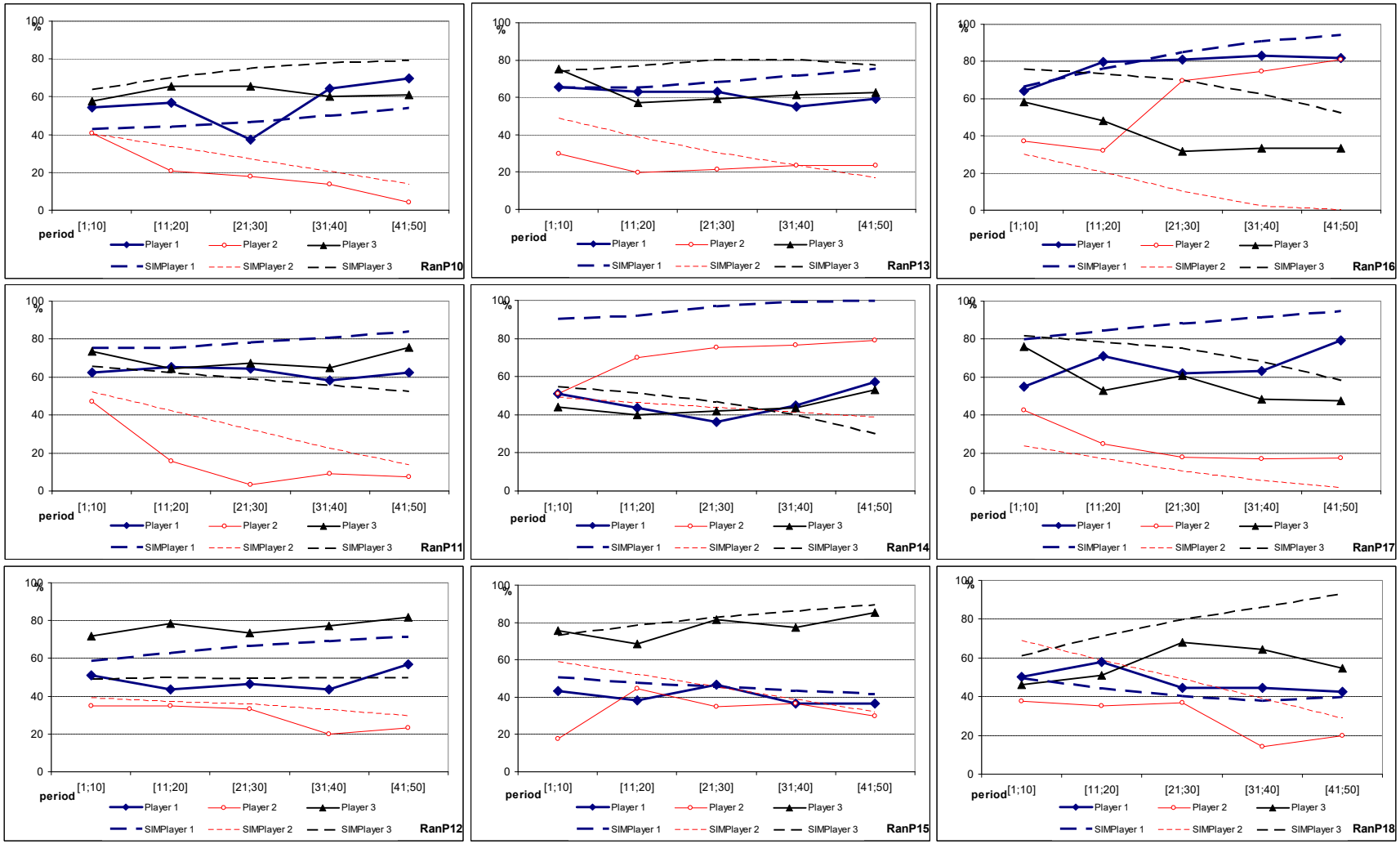


Figure 6.2 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay treatment: average behavior strategies, probability of playing R, over 10 period

In order to present a more competitive benchmark than chance, we also conduct the simulation of reinforcement dynamics (Erev and Roth 1998).[11] As with the previous two simulations we start at the initial choices and matchings of the 9-subjects group. The resulting dynamics are reinforced in the following manner;

$$\widetilde{p}_{it} = \frac{\alpha_{it}^{R}}{\alpha_{it}^{L} + \alpha_{it}^{R}}, \qquad where$$

$$\begin{aligned} \alpha_{it}^{R} &= \alpha_{it-1}^{R} + e_{it-1}^{R}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1}) \\ \alpha_{it}^{L} &= \alpha_{it-1}^{L} + e_{it-1}^{L}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1}) \end{aligned}, \qquad (5)$$

$e_{it-1}^{j}(\widetilde{p}_{it-1}, \widetilde{p}_{-it-1})$ is the expected payoff when playing $j = L, R$ conditionally of the other players $-i$ employing their simulated behavior strategy, and α_{i1}^{R} is the observed first behavior strategy of subject i multiplied by 100. Again, we measure the average squared deviation of each subject's reinforcement trajectory from the observed trajectory over each 10-period interval and sum the deviations over the nine players and 5 time intervals. The simulation is conducted for each of the 18 independent sessions, and repeated 1,000 times. Comparing the average mean-squared error of the random trajectories with the error of the reinforcement trajectories, we find that in 13 of 18 sessions the latter is smaller than the former. The probability that 13 out of 18 or more successes would be drawn by chance is as low as .015. Thus, reinforcement learning also predicts the outcomes of the choices better than chance. However, compared to the impulse response trajectories the mean squared error is significantly larger. The same 14 sessions of 18 sessions that are better predicted with impulse response than by chance are also better predicted than with reinforcement dynamics. Hence, impulse response predicts the observed dynamics better than reinforcement learning, too.

We note that the reinforcement dynamics do not necessarily converge on an equilibrium. As shown above in Figure 5, equilibrium adjustments may be non-monotonic with the impulse response model. In contrast, reinforcement trajectories are monotonic in our case. Each reinforcement trajectory converges towards the upper or lower boundary of

[11] Our adaptive approach of impulse response is outcome oriented, and is parameter free. So 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.

behavior strategies. Once a boundary is reached the trajectories settles there. Typically, depending on the starting point, the reinforcement trajectories move fast in early periods and thereafter need a very long time (many thousands simulation periods) to converge on a rest point. Depending on the starting point any strategy profile of extreme behavior strategies can institute a final rest point.

6 Alternative theories of non-equilibrium behavior

Our data show non-equilibrium behavior in the Game of Selten's Horse. In other games, quantal response dynamics (McKelvey and Palfrey 1995), the level-k model (Nagel 1995, Crawford 2013) and Pareto efficiency have been useful to illuminate non-equilibrium behavior (e.g., Garcia-Pola et al. 2020). The former two approaches apply best-response dynamics, and the latter is a traditional approach that evaluates the efficiency of the outcomes.

6.1. Quantal response equilibrium

Similarly to trembling hand perfection, the quantal response approach (McKelvey and Palfrey 1995) allows that players 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. Differently, however, trembling-hand perfection selects the equilibrium on the basis of robustness against errors, whereas the quantal response trajectory selects the equilibrium profile by reducing errors until they vanish.

The following set of equations shows the logit quantal-response functions for the Game of Selten's Horse of the noise parameter λ^{-1}; λ is assumed to be close to zero for inexperienced players and large for experienced subjects.

$$\begin{aligned} p_1(p_2, p_3, \lambda) &= \frac{1}{1+\exp(-\lambda(p_2 + p_3 - 4p_2p_3))} \\ p_2(p_1, p_3, \lambda) &= \frac{1}{1+\exp(-\lambda p_1(1-4p_3))}, \\ p_3(p_1, p_2, \lambda) &= \frac{1}{1+\exp(-\lambda(2-3p_1+2p_1p_2))} \end{aligned} \tag{6}$$

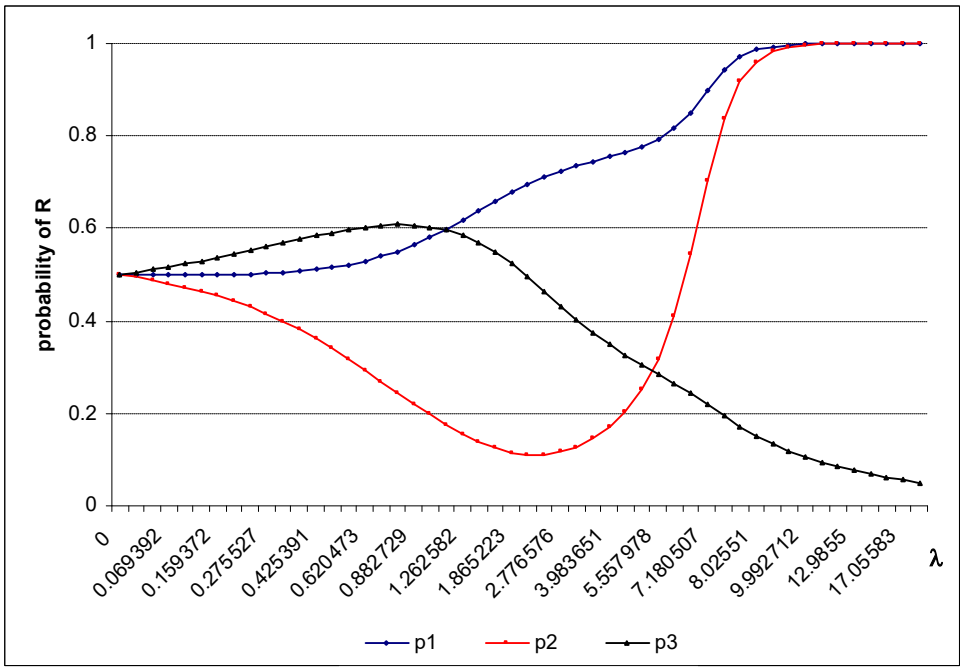


Figure 7. Principal branch of quantal response correspondence

With perfect noise, $\lambda = 0$, the proposed strategy profile is (0.5, 0.5, 0.5) and when noise vanishes, as $\lambda \to \infty$, the quantal response correspondence selects the perfect equilibrium profile (1, 1, 0). Figure 7 displays the quantal response curves, which are attracted to the trembling-hand perfect equilibrium.[12] We note the similarity to Figure 5; the quantal response curves look like a smooth version of the impulse response trajectories. Before the quantal response curves reach the perfect equilibrium set at $\lambda \geq 18$, they describe non-equilibrium behavior.

The initial choices of subject types 1 and 2 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. The observed initial cumulative distribution is depicted in Figure A1.2 of the appendix.

Similar to Capra et al. (1999) and Goeree and Holt (1999), we estimate $\lambda = 0.31$, s.d. = .029, applying the MLE to the overall data. For the first period we have an estimate of 0, s.d. = .10. For the first (last) ten periods, our estimates of λ are 0.03 (0.445), with s.d. of

[12] We made use of the Gambit software (McKelvey et al. 2014) to compute the principal curve of the logit quantal response correspondence.

0.031 (0.028). These estimates indicate a reduction of noise and a movement towards the trembling-hand perfect equilibrium over time in our experiment.

6.2 Level-k non-equilibrium model

The level-k model of cognitive reasoning 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 (Garcia-Pola et al. 2020), level-k reasoning converges to common knowledge of rationality as $k \rightarrow \infty$. In the Game of Selten's Horse, however, level-k responses do not converge as the level of reasoning is increased, but start to cycle instead. 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.

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
…	…	…	…	…	…	…	…	…	…

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 with k = {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) closest to the one we observe in our data. Table 4 shows the level-k mixtures we identify overall and for each treatment.[13]

Table 4. Squared error minimizing level-k mixtures

	Level 0	Level 1	Level 2	Level 3
Average Pay	0.550	0.450	0.000	0.000
Random Pay	0.800	0.150	0.050	0.000
Overall	0.650	0.350	0.000	0.000

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 vast majority of the others reveals only a level-1 reasoning. However, as explained in the next subsection, 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 drive Pareto efficient outcomes by choosing a behavior strategy close to the 50-50 mixture (i.e., $p_1 = 0.5$). This form of cooperation probably entails a high level of reasoning, even though it is identified as level-0 behavior in the level-k model.

[13] 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.

6.3 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. Collective rationality pits against the individual risk of trembling in decision making. Pareto efficiency is defined as any allocation from which no player can get better off without making another player worse off. The necessary condition to reach the Pareto efficient allocations in the Game of Selten's Horse is player 3 chooses R. 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. Note that in each of our experiments more than half of the outcomes are Pareto efficient. The Random Pay treatment has a lower frequency of Pareto efficient outcomes than the Average Pay treatment, although the effect is not significant; player 3 receives nothing if the final outcome is z_4, therefore is prone to regret her decision $p_3 > 0$, and to reduce the probability of playing R.

Player 3 chooses R in a best-response manner when node x_1 (following player 1's choice left) is at least as likely as node x_2 (following right and left choices of players 1 and 2, respectively). If, for instance, player 1 chooses $p_1 \leq 0.5$, the best response of player 3 is to choose $p_3 = 1$. Player 2's best response is then to choose $p_2 = 0$.

The way the game is played depends crucially on the strategy of player 1, who thus has a first-mover advantage. Given the best responses to player 1, any strategy $p_1 \in (0, 0.5)$ has a higher expected payoff than player 1 can achieve in any equilibrium. Despite the fact that $p_1 = 0.5$ implies the best responses $p_2 = 0$ and $p_3 = 1$, it is no equilibrium strategy, since player 1 does not play a best response to the strategy profile of the other players.[14] As shown in Table 1, the majority of choices (i.e., 53.9%) involves the interval $p_1 \leqslant 0.5$, and the average choice $p_1 = 0.480$ is also contained in that interval. At the same time, the modal choices of player types 2 and 3 confirm the indicated pure best response strategies to this play.[15]

[14] It is an attractive strategy because it is sustainable, whereas the outcome (4,4,0), in which player 1 plays the best response to the described strategy profile of the other players, is unsustainable. Assuming rationality of all players, it should be a focal (non-equilibrium) play.

[15] Particularly, player type 2 deviates from her equilibrium play. In any equilibrium she should play R with a probability of at least 0.25. The data suggest that the majority of type 2 players violate that prediction. The triggering point of such behavior is that type 3 players choose right with a probability of above 0.25, thus making it type 2 players best response to choose L over R.

7 Robustness test: Long-run behavior

The impulse response adjustment dynamics presented in section 5 and also the quantal response dynamics in section 6.1 can entertain speculations about the long-run behavior in experiments in view of trembling-hand perfection. The view-point of level-k or Pareto efficiency, on the other hand, would suggest a continuation of non-equilibrium behavior or a move towards the imperfect equilibrium set.

This question needs to be addressed; will subjects' behavior be attracted to the perfect equilibrium in the long-run game? We analyze the question in this section. To check convergence on any equilibrium, we conducted longer sessions than in the first study. A two-hour long session would allow for up to 250 periods of interaction. Like in the first experiment, subjects were students of the University of Magdeburg, and every student participated in one cohort. Depending on their pace, some cohorts were interacting faster and others were interacting slower. Since we stopped the experiment at the sooner, after 2 hours or after 250 periods, the sessions ended after a different number of periods.

To give the theory an excellent chance to succeed in the experiment, we considered both a strangers' setting and a partners' setting. In section 7.1, we report on the former one, and we report on the latter one in section 7.2.

7.1 Experiment 2: long-run strangers' experiment

The experimental design was almost identical to the first study, only that the subjects were invited for a two-hour session, and the numbers of periods varied between 118 and 235 depending of the pace of cohorts instead of 50 periods for everyone. Just as in the first experiment, we had 9 cohorts of nine subjects of each, the Random Pay treatment and the Average Pay treatment. Figure A2.1 and Table A2.1 in the Appendix preview the outcomes.

The relative frequencies of the outcomes of the long-run strangers' experiment are surprisingly similar to the ones reported for the first experiment with only 50 periods. Comparing the overall relative frequencies for the experiments, we see almost no treatment effect of the long-run average behavior regarding the use of strategy (as in Table A2.1) or outcomes; the p-values of the Mann-Whitney test comparing the first experiment with the long-run strangers' experiment almost always exceed the 10-percent level, but once.[16]

[16] Only the first-type choice of playing right is (weakly) significantly different between the treatments; the p-value of the Mann-Whitney test is 0.0935 when we compare the relative frequency of playing right with probability 1 across treatments. If this effect suggests anything, then that over time the pure right choice is less frequently observed in the long-run, contrary to what the perfect equilibrium proposes. The relative frequency of

Observation 9: The long-run behavior in the strangers' setting does not converge on the perfect equilibrium.

The only indication of strategy adjustment that differs between the first experiment and the long-run strangers' experiment indicates rather a move away than towards the perfect equilibrium with longer sessions. Although the comparison across treatments does not show an impact on the outcomes, we find such an impact in within-subjects comparison. For the first 50 periods, the relative frequency of the perfect-equilibrium outcome (1, 1, 1) is 0.1887, but it is only 0.1522 for the last 50 periods. The difference is weakly significant as the two-tailed Wilcoxon signed ranks test confirms; the p-value is 0.0936.

Figures A2.1 and A2.2 display the average choice of the different player types together with the simulated impulse response trajectories by period interval. As above, each impulse-response trajectory initiates at the subjects' first choice. The overall impression is different from the first experiment, as the impulse response trajectories rarely follow the observed choice over the longer time horizon. A typical choice trajectory starts at an intermediate probability of playing right, then moves to a higher or a smaller probability, where it remains relatively constant at no extreme probability level. The simulated impulse response trajectory, in contrast, follows its complex trajectory towards the pure trembling-hand-perfect equilibrium. As in the first experiment, player type 1 and particularly type 2 usually choose to play left rather than right and player type 3 plays right rather than left.[17] This kind of non-equilibrium behavior makes convergence on any equilibrium unlikely.

For the long-run strangers' experiment, we obtain the following estimates of the logit quantal response model. We estimate $\lambda = 1$, s.d. = .133, applying the MLE on the overall data. For the first (last) ten periods, we estimate $\lambda = 0.19$, s.d. = .109, ($\lambda = 1.07$, s.d. = .137), and for the first period $\lambda = 0$, s.d. = .129. Compared to the first experiment, the noise level seems reduced in the long-run experiment, particularly for the later periods.

Running the same level-k analysis for our long-run strangers' experiment as we have presented for the data of the first experiment, we identify the following level-k mixtures (level-0, level-1, level-2, level-3) minimizing the squared error of the predicted to observed behavior strategy vectors: (.60, .40, .00, .00) overall, (.65, .30, .05, .00) for the first 50 rounds,

---

the outcomes was not different across first and second experiment. We tested it for all periods of the long-run experiment as well as for the first 50 periods and the last 50 periods separately.

[17] In four cohorts (AvgP21, AvgP23, RanP29, RanP30) player 1 plays rather right and player 3 plays rather left.

and (.60, .40, .00, .00) for the last 50 rounds. As in the first experiment, the identified levels of reasoning in the long-run sessions seem very low, with substantially more than 50 percent of the subjects classified as level-0 types. Again, we conjecture that the detected low levels of reasoning do not reflect the subjects' actual capabilities to analyze the game strategically and find the best response, but are due to the cooperative play that is especially driven by player 1 choosing an almost perfectly mixed strategy to enforce the Pareto efficient outcome of the game.

7.2 Experiment 3: long-run partners' experiment

Finally, we conducted long-run sessions of the Game of Selten's Horse in a partners' setting. Three subjects interacted repeatedly in the two-hour long session, with three cohorts interacting on one server at the same pace. The number of periods that were played varied between 111 and 158 depending on the pace of the groups' interaction. The experiment involved 9 cohorts of each, the Random Pay treatment and the Average Pay treatment.

The partners' setting is interesting as it makes equilibrium selection relatively easy. The partners' experiment can inform us about coordination problems that may arise in the strangers' experiment. In the partners' setting, however, there exist group incentives for cooperative play to capture the continuation payoff from cooperation.

Figures A2.2, A3.3 and A3.4, and Table A2.2 in the appendix and Figure 7 show the outcomes and the average frequencies of strategy usage. Relative to the other experiments, we can state for the partners' experiment the following.

Observation 10: Subjects use pure strategies more frequently. The imperfect equilibrium outcome and the Pareto efficient outcome are more frequently achieved in the partners' setting than in the strangers' setting.

The average share of pure strategies is 0.660 in the partners' setting, whereas it is 0.370 in the long-run strangers' setting (and 0.433 in the first experiment). The difference is significant; a p-value of 0.007 (0.021) according to the Mann-Whitney test. The average relative frequency of obtaining the imperfect (perfect) equilibrium outcome is 0.734 (0.099) in the partners' experiment, whereas it is 0.409 (0.156) in the long-run strangers' experiment and 0.368 (0.169) in the first experiment. The differences are significant; according to the Mann-Whitney test the p-value is 0.001 (0.002) with respect to the strangers' experiment and 0.000 (0.004) with respect to the first experiment. The share of Pareto-efficient outcomes is

0.814 in the partners' experiment and 0.582 and 0.606 in the first and in the long-run strangers' experiment; the difference is significant, $p < 0.003$. The differences between the long-run strangers' and the first experiment are not significant in any of these comparisons at the 10-percent level.

Interestingly in the Partners' experiment we observe in one cohort (RanP51) the perfect equilibrium play repeatedly for almost 100 periods. Suddenly, however, player 1 switches from right to left and player 3 from left to right to continue playing this way for the following 40 periods until the end of the experiment. This pattern suggests that the trembling-hand perfect equilibrium can not be a stable outcome in the repeated setting. Similarly to social dilemma games, the benefits from collaboration stand out.

We remark another cohort in the experiment (RanP54) that shows how player 1 can benefit from the first mover advantage. In that cohort, the player 1 chooses strategy $p_1 = 0.5$ in all periods and triggers thus the best responses from players 2 and 3. The outcome is in 96% of periods Pareto efficient. That subject of player type 1 achieved the highest payoff even though other cohorts managed to finish more periods of play in the provided time.

For the long-run partners' experiment, we obtain the following estimates of the logit quantal response model. For the long-run partners experiment, we estimate $\lambda = 1.2166$, s.d. = .2432, applying the MLE on the overall data. For the first (last) 10 periods the estimates are 1.65, s.d. = .2527 (1.17, s.d. = .3613).

7.3 Dynamics

We find that the convergence patterns are different in the partners' experiment than in the strangers' experiment. Figure 7 displays the trajectories for the first 50 and last 50 periods. Player types 2 behave similarly in both treatments, player types 1 and 3 play more frequently pure strategies in the partners' experiment. Hence, in the partners' experiment the behavior strategies of these players are close to the imperfect equilibrium play. That explains why the outcome frequency of z_2 is higher and the outcome frequency of z_4 is lower in the partners' than in the strangers' setting.

Direction learning is supported in the long-run experiments by the choices of the majority of subjects, see Tables A3.1 and A3.2 in the appendix. The impulse response trajectories (see Figures A3.1-4), however, seem to provide no good fit for the long term behavior in the Game of Selten's Horse. The impulse response is attracted towards the trembling-hand perfect equilibrium in the long-run, but the observed behavior is not. It seems fair to say that it describes well the behavioral dynamics of our experimental sessions for 50

periods. In the long run, adjustments towards the Pareto-efficient outcomes explain the deviations of the observed behavior to the impulse response trajectories. The group decision process is frequently quite stable over time and seems to converge fast, particularly in the long-run partners' experiment.

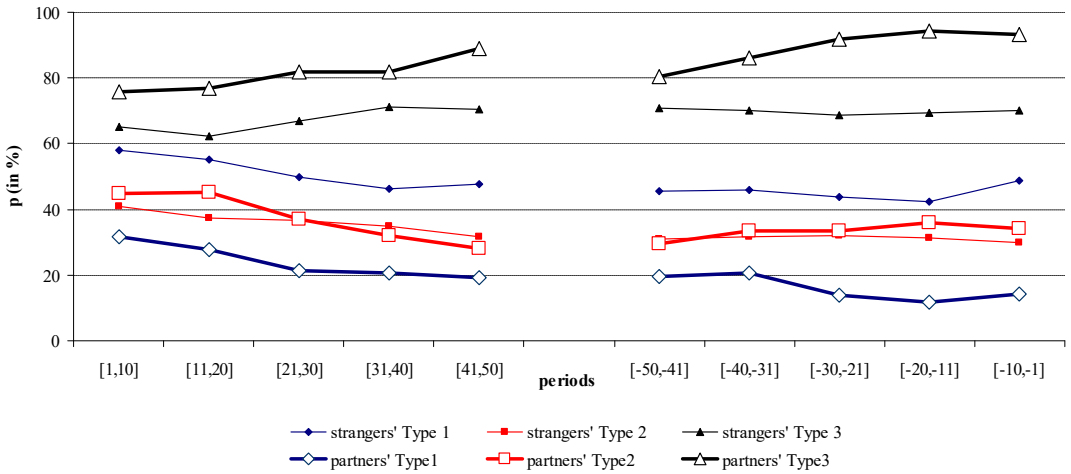


Figure 7 Average behavior strategies of player types 1 to 3 in the long-run experiments
Each dot represents the ten-period average probability of playing R for the first and last fifty periods. Thin lines/filled dots (thick lines/empty dots) show the trajectories of the strangers' (partners') experiment.

8 Conclusions

We report experimental data on the Game of Selten's Horse (Selten 1975) and suggest impulse response trajectories as an appropriate simulation approach for explaining the short-term and intermediate behavior in the game. Impulse response is a one-step, deterministic simulation application of best-response dynamics interrelated with learning direction theory (Selten and Stoecker 1986), which receives empirical support in our experiment as well as in many other experiments. In contrast to other best-response learning applications (as e.g., Cournot learning), the impulse response trajectories account for the inertia in human interaction and predict how dynamic adjustments approach the equilibrium. In future research, it will be interesting to see how well the behavioral dynamics in other game environments are described by the impulse response trajectories.

For the long run, the impulse response trajectories suggest that the dynamic adjustments of strategy choices in the Game of Selten's Horse will drive behavior to the trembling-hand perfect equilibrium set. The logit quantal response equilibrium (McKelvey and Palfrey 1995) also suggests the selection of a trembling-hand perfect equilibrium as the

point of convergence when the noise vanishes. Notwithstanding, as our experimental data show, the trembling-hand perfect equilibrium is unattractive to subjects playing the Game of Selten's Horse.

Our experimental implementation allows the play of all equilibrium strategy profiles. However, equilibrium profiles are rarely observed in the experiments and the relative frequency of the trembling-hand perfect equilibrium outcomes declines with repetition.[18] Although the unreasonable imperfect equilibrium outcomes as suggested in Selten (1975) are obtained more frequently than the trembling-hand perfect equilibrium outcomes, we also find limited support for the imperfect equilibrium in our data. The most frequently observed strategy profile in our data 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. Even if player 1 could still increase his payoff by playing the trembling-hand perfect strategy, he obtains three times the payoff from playing his imperfect equilibrium strategy than he would make in the trembling-hand perfect equilibrium. To obtain large continued payoff from cooperation for all players, the third player must be hooked on the right strategy.

The observed play is characterized by non-equilibrium behavior, which can be modeled by level-k reasoning (Crawford 2013) as we have shown. Nonetheless, it is not irrational play; Pareto efficient outcomes are reached frequently. Subjects usually achieve a higher payoff than the one predicted by trembling-hand perfection. Trembling-hand perfection seems a rational way of play only for a short period of time in a population with no trust and forgiveness, or as a potential threat point. For the long run, collective rationality seems to favor payoff-richer non-equilibrium profiles in the Game of Selten's Horse.

[18] In independent research, Berninghaus, Güth and Li (2012) studied 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. Despite the differences to our design their data also give little support to the perfect equilibrium. In line with our observed dynamics, Berlinghaus et al. (2012) wonder whether the perfect equilibrium may have a better chance of emerging as a stable and frequent outcome in a repeated game.

References

Binmore, K., 1987, Modeling Rational Players: Part I, Economics and Philosophy 3(2), 179-214.

Berninghaus, S., Güth, W. and K.K. Li, 2012, Approximate truth of perfectness: An experimental test, Karlsruhe Institute of Technology, Working Paper series in economics, No. 41.

Camerer, C., 2003, Behavioral game theory: Experiments in strategic interaction. Princeton University Press.

Capra, C. M., Goeree, J. K., Gomez, R., & Holt, C. A. (1999). Anomalous behavior in a traveler's dilemma?. American Economic Review, 89(3), 678-690.

Cheung, Y.-W., and D. Friedman, 1997, Individual learning in normal form games: Some laboratory results. Games and Economic Behavior 19(1), 46-76.

Chmura, T., Goerg, S.J. and R. Selten, 2012, Learning in experimental 2× 2 games. Games and Economic Behavior 76(1), 44-73.

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

Costa-Gomes, M.A., and V.P. Crawford, 2006, Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. American Economic Review 96(5), 1737-1768.

Crawford, V.P., 2013, Boundedly Rational versus Optimization-Based Models. Journal of Economic Literature 51(2), 512–527.

Crawford, V.P., Costa-Gomes, M.A. and N. Iriberri, 2013, Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications. Journal of Economic Literature, 51(1), 5–-62.

Erev, I., and A.E. Roth, 1998, Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. The American Economic Review 88(4), 848-881.

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

Friedman, D., and R. Oprea, 2012, A Continuous Dilemma. American Economic Review 102(1), 337–363.

García-Pola, B., Iriberri, N., and J. Kovářík, 2020. Non-equilibrium play in centipede games. Games and Economic Behavior, 120, 391-433.

Goeree, J.K. and Holt, C.A., 1999. Stochastic game theory: For playing games, not just for doing theory. Proceedings of the National Academy of sciences, 96(19), pp.10564-10567.

Goerg, S., Neugebauer, T. and A. Sadrieh, 2016. Impulse response dynamics in weakest link games. German Economic Review 17 (3): 284-297.

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

McKelvey, R. D., and T. R. Palfrey, 1995. Quantal response equilibria for normal form games. Games and economic behavior, 10(1), 6-38.

Nagel, R., 1995, Unraveling in guessing games: An experimental study. The American economic review, 85(5), 1313-1326.

Neugebauer, T., and R. Selten, 2006, Individual behavior of first-price auctions: The importance of information feedback in computerized experimental markets. Games and Economic Behavior 54(1), 183-204.

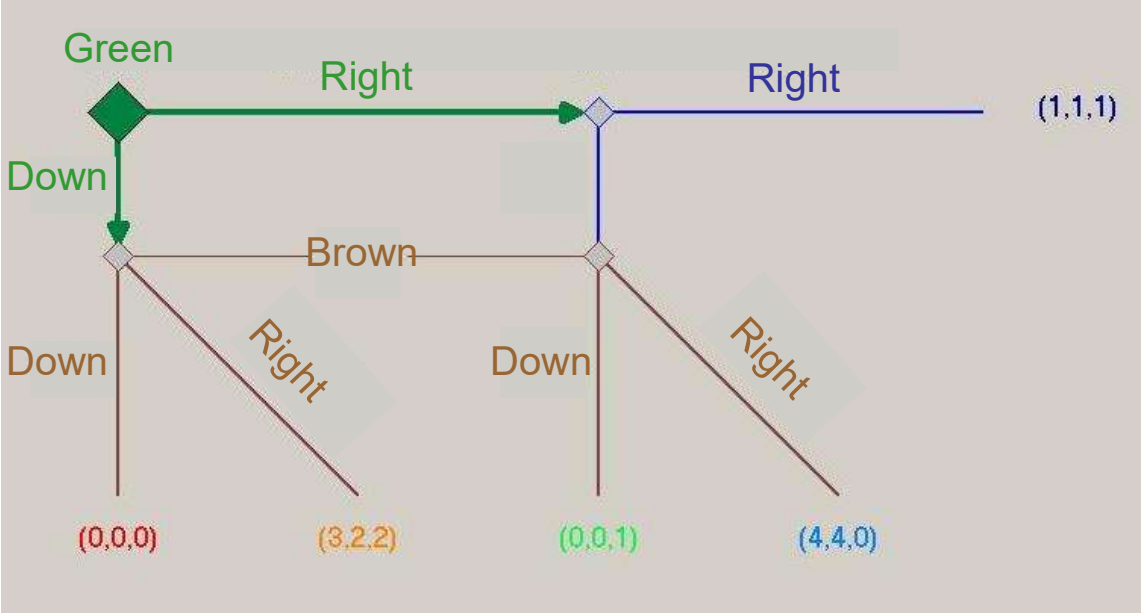
Nyarko, Y., and A. Schotter, 2002, An experimental study of belief learning using elicited beliefs. Econometrica 70(3), 971-1005.

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

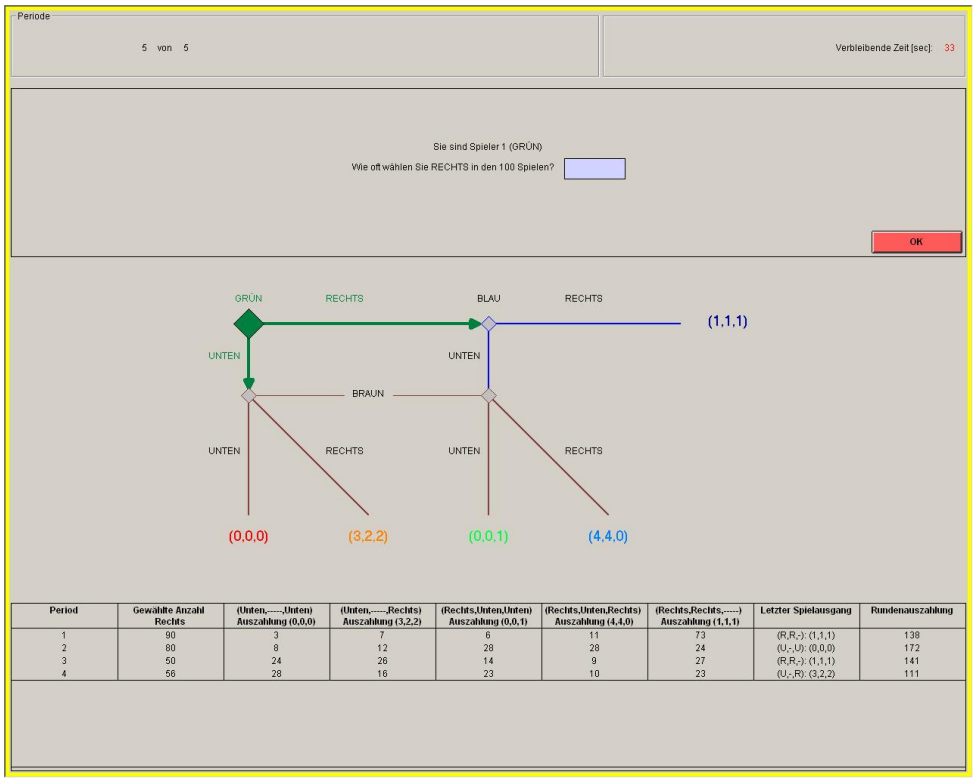
Selten, R., 1965, Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit. Zeitschrift für die gesamte Staatswissenschaft 121, 301-327 and 667-689.
Selten, R., 1973, A simple model of imperfect competition, where 4 are few and 6 are many. International Journal of Game Theory 2, 141-201.
Selten R. 1975, Reexamination of the perfectness concept for equilibrium points in extensive games, International Journal of Game Theory 4(1), 25-55.
Selten, R., 2004, Learning direction theory and impulse balance equilibrium. D. Friedman, A. Cassar (Eds.), Economics Lab—An Intensive Course in Experimental Economics, NY: Routledge, 133–140.
Selten, R., K. Abbink, and R. Cox, 2005, Learning direction theory and the winner's curse. Experimental Economics 8(1), 5-20.
Selten, R., and J. Buchta, 1999, Experimental sealed bid first price auctions with directly observed bid functions. In: A. Rapoport, D.V. Budescu, I. Erev, and R. Zwick (eds.), Games and Human Behavior: Essays in Honor of Amnon Rapoport, 101-116.
Selten, Reinhard, and Thorsten Chmura. 2008. Stationary Concepts for Experimental 2x2-Games. American Economic Review, 98 (3): 938-66.
Selten, Reinhard, Thorsten Chmura, and Sebastian J. Goerg. 2011. Stationary Concepts for Experimental 2 X 2 Games: Reply. American Economic Review, 101 (2): 1041-44.
Selten, R., and R. Stoecker, 1986. End behavior in sequences of finite prisoners' dilemma supergames: A learning theory approach. J. Econ. Behav. Organ. 7, 47–70.

Appendix. Instructions

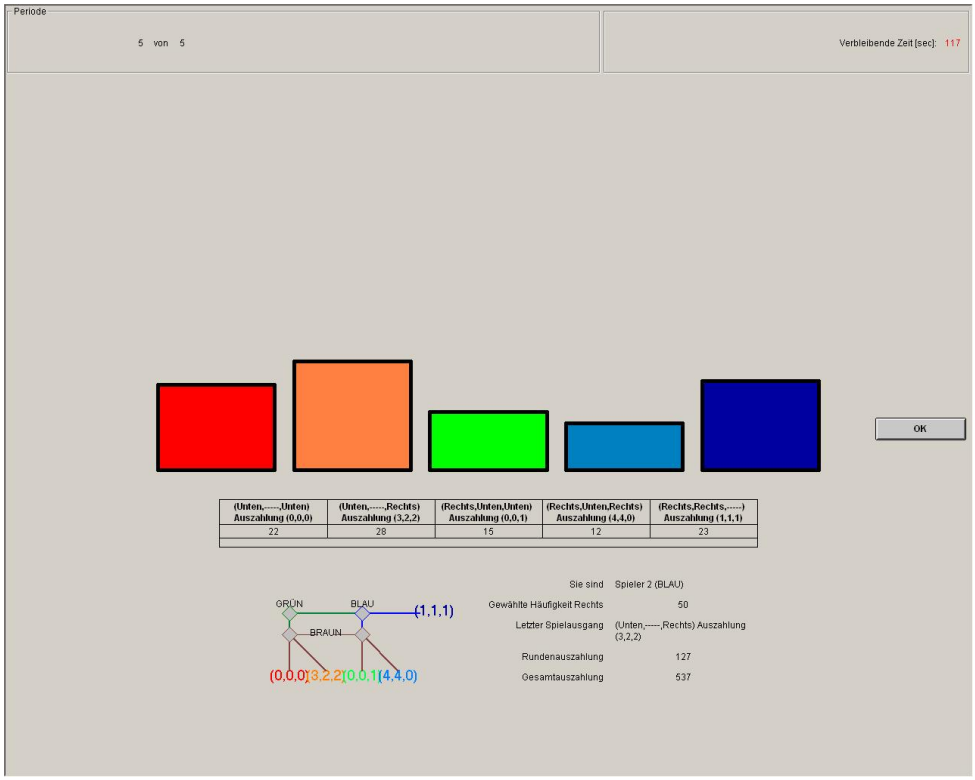
- This experiment involves two sequences of 50 rounds.
- In each round, you make decisions for 100 three-person games, which are played simultaneously.
- In every round you are matched and play with two other participants in the experiment. The matching in each round are random. The probability to play 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 treatment:] You earn your average payoff in the 100 games of the round.
- [random pay treatment:] 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.



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



(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

Appendix. Tables

Table A1. Mean squared errors of simulation and choice trajectories for each cohort

cohort	impulse response	random	reinforcement
1	284	953	935
2	447	537	488
3	161	374	270
4	135	208	140
5	515	1392	1302
6	155	355	243
7	968	574	664
8	71	204	138
9	1308	851	997
10	244	692	501
11	336	355	348
12	725	515	598
13	972	834	912
14	155	399	298
15	155	400	298
16	317	334	350
17	477	704	613
18	1175	1230	1221

The mean squared errors are computed by summing the squared deviations of simulation and choice and taking averages

Table A2.1 Cumulative distribution of behavior strategies in the long-run strangers' experiment

		p=0	[0,.25]	[0,.33]	[0,.50]	[0,.75]	[0,.99]	[0,1]
player 1	AvgPay	**0.192**	0.356	0.407	0.563	0.756	0.892	1
	RandomPay	**0.237**	0.342	0.371	0.529	0.692	0.885	1
	Overall	**0.229**	0.361	0.398	0.542	0.728	0.890	1
player 2	AvgPay	**0.321**	0.597	0.641	0.797	0.922	0.973	1
	RandomPay	**0.294**	0.467	0.496	0.682	0.808	0.882	1
	Overall	**0.306**	0.538	0.574	0.747	0.867	0.927	1
player 3	AvgPay	0.095	0.150	0.187	0.271	0.378	0.581	**1**
	RandomPay	0.075	0.168	0.191	0.304	0.506	**0.785**	1
	Overall	0.086	0.160	0.191	0.283	0.427	0.668	**1**

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

Table A2.2 Cumulative distribution of behavior strategies in the long-run partners' experiment

		p=0	[0,.25]	[0,.33]	[0,.50]	[0,.75]	[0,.99]	[0,1]
player 1	AvgPay	**0.737**	0.838	0.850	0.877	0.956	0.991	1
	RandomPay	**0.431**	0.590	0.623	0.860	0.878	0.900	1
	Overall	**0.572**	0.703	0.727	0.871	0.917	0.946	1
player 2	AvgPay	**0.564**	0.600	0.610	0.663	0.712	0.867	1
	RandomPay	**0.358**	0.593	0.607	0.705	0.752	0.842	1
	Overall	**0.461**	0.604	0.615	0.687	0.731	0.855	1
player 3	AvgPay	0.046	0.062	0.070	0.085	0.107	0.135	**1**
	RandomPay	0.100	0.127	0.143	0.177	0.234	0.547	**1**
	Overall	0.073	0.094	0.107	0.131	0.170	0.341	**1**

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

Table A3.1 Direction learning in the long-run strangers' experiment

Strangers		Dir learning surplus subjects	Dir learning surplus + no change subjects
type 1	=	3	0
	<	9	0
	>	42	54
type 2	=	3	0
	<	11	2
	>	40	52
type 3	=	1	0
	<	2	0
	>	51	54
Total	=	7	1
	<	22	3
	>	133 (82%)	160 (99%)

=/< /> indicate the number of subjects that respond as/less/more frequently in the direction predicted by as/than/than opposing to direction learning theory

Table A3.2 Direction learning in the long-run partners' experiment

Partners		Dir learning surplus subjects	Dir learning surplus + no change subjects
type 1	=	4	0
	<	0	0
	>	14	18
type 2	=	0	0
	<	5	1
	>	13	17
type 3	=	6	0
	<	6	1
	>	6	17
Total	=	10	0
	<	11	2
	>	33 (61%)	52 (96%)

=/< /> indicate the number of subjects that respond as/less/more frequently in the direction predicted by as/than/than opposing to direction learning theory

Appendix. Figures

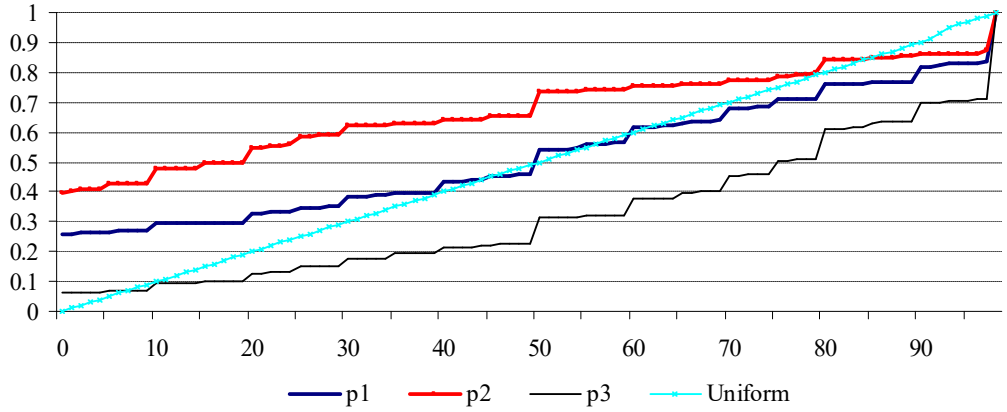


Figure A1.1 Cumulative choice distribution organized by type

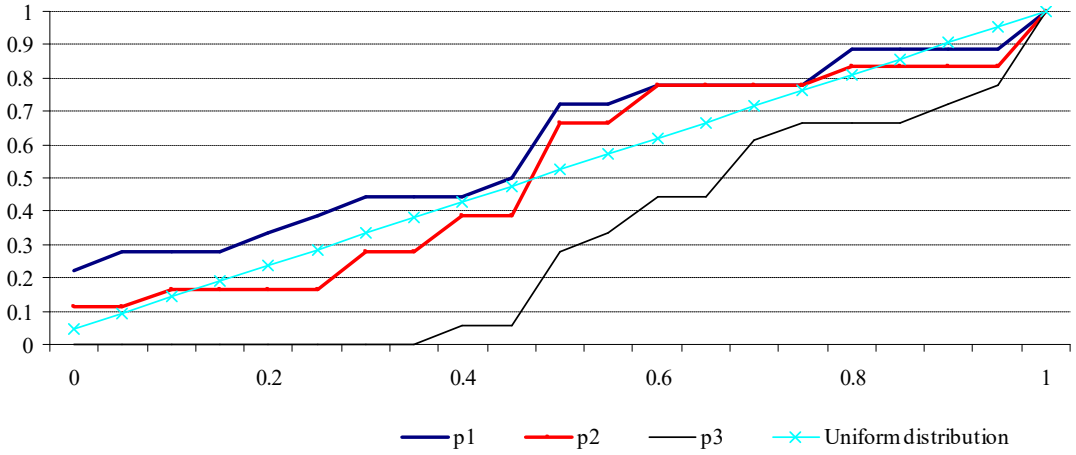


Figure A1.2 Cumulative distribution of subjects' first period choices organized by type

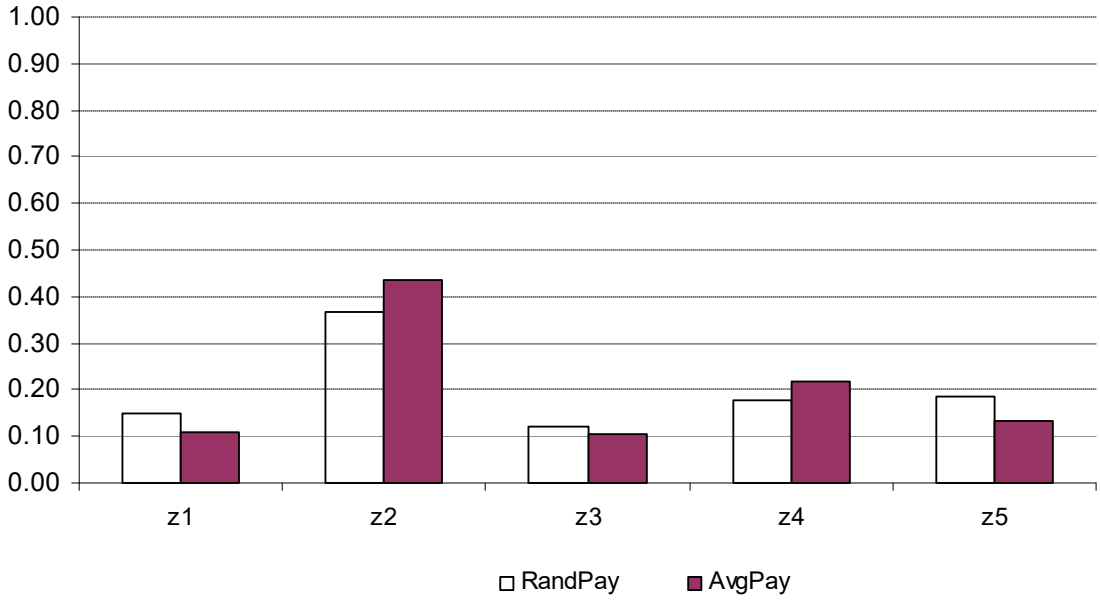


Figure A2.1 Average outcomes in the long-run strangers' experiment

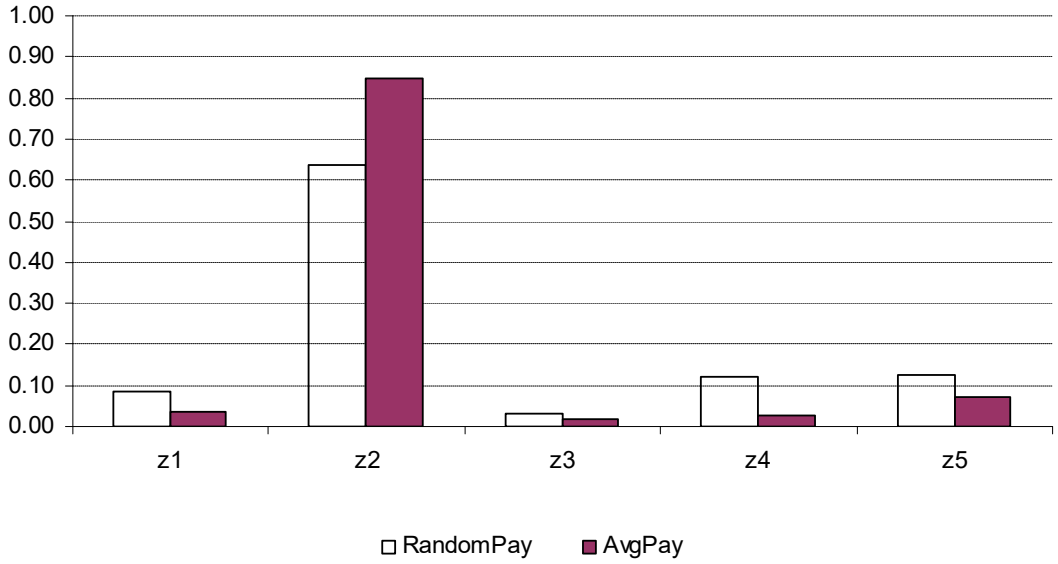


Figure A2.2 Average outcomes in the long-run partners' experiment

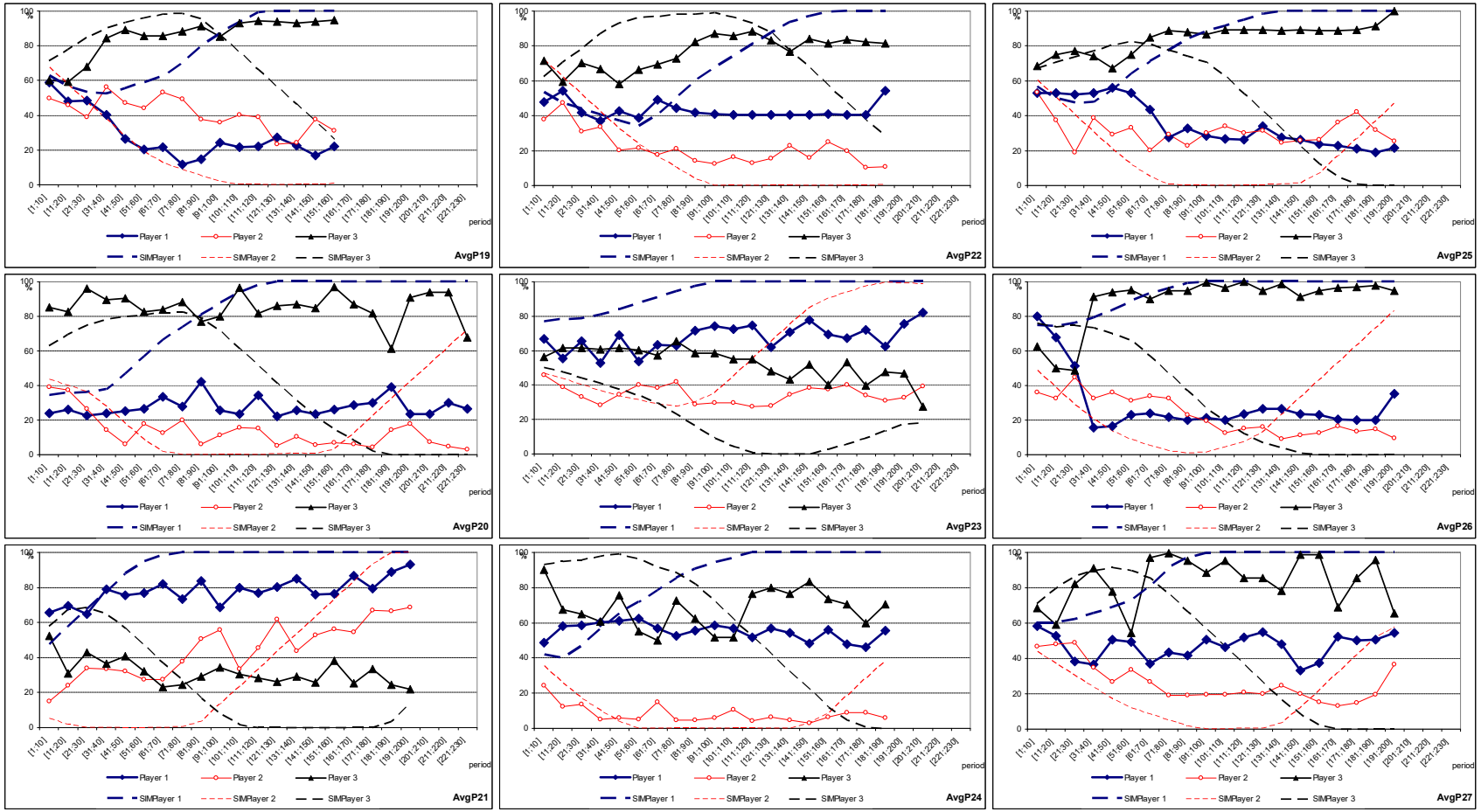


Figure A3.1 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay treatment of the long-run strangers' experiment: average behavior strategies, probability of playing R over 10 periods

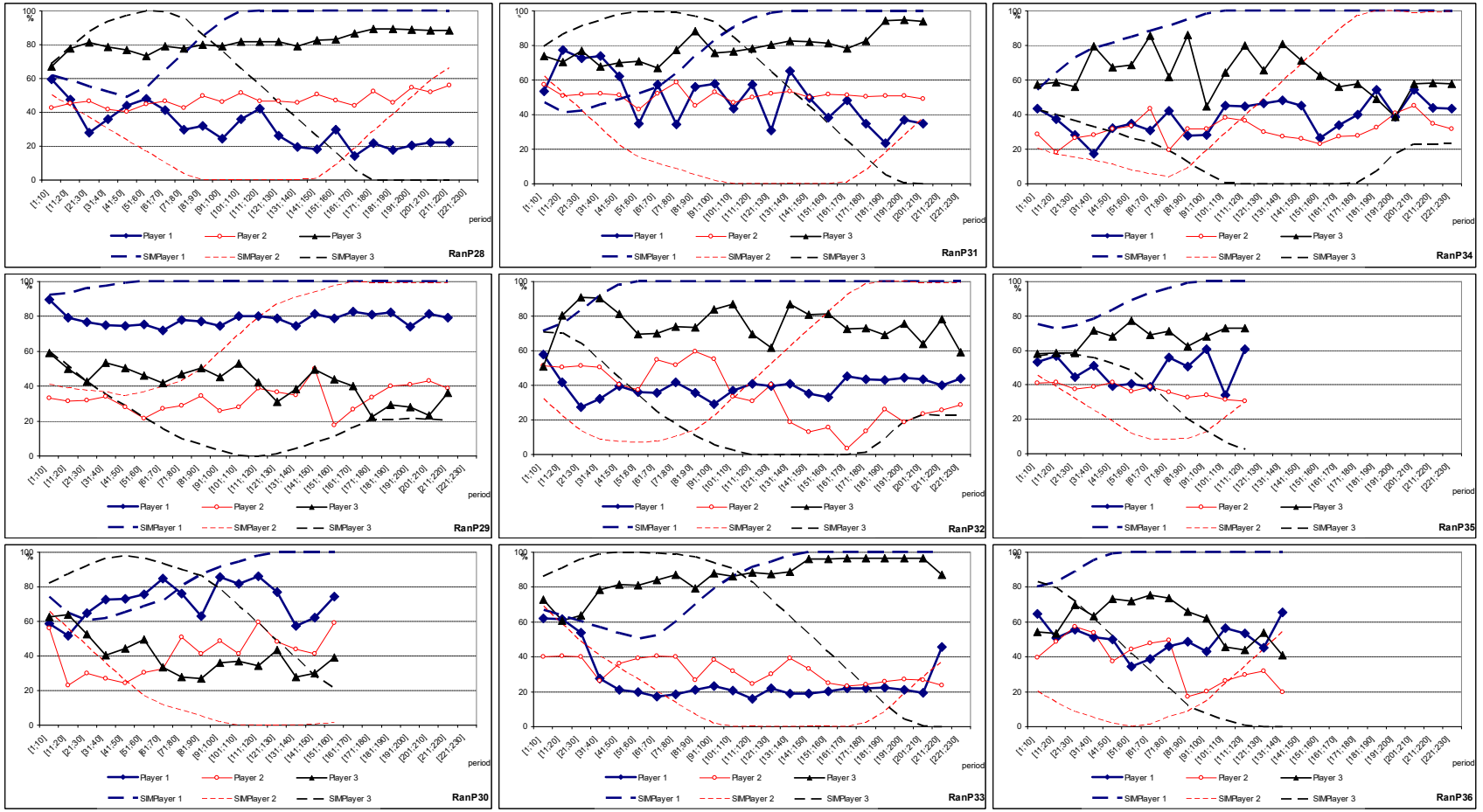


Figure A3.2 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay treatment of the long-run strangers' experiment: average behavior strategies, probability of playing R over 10 periods

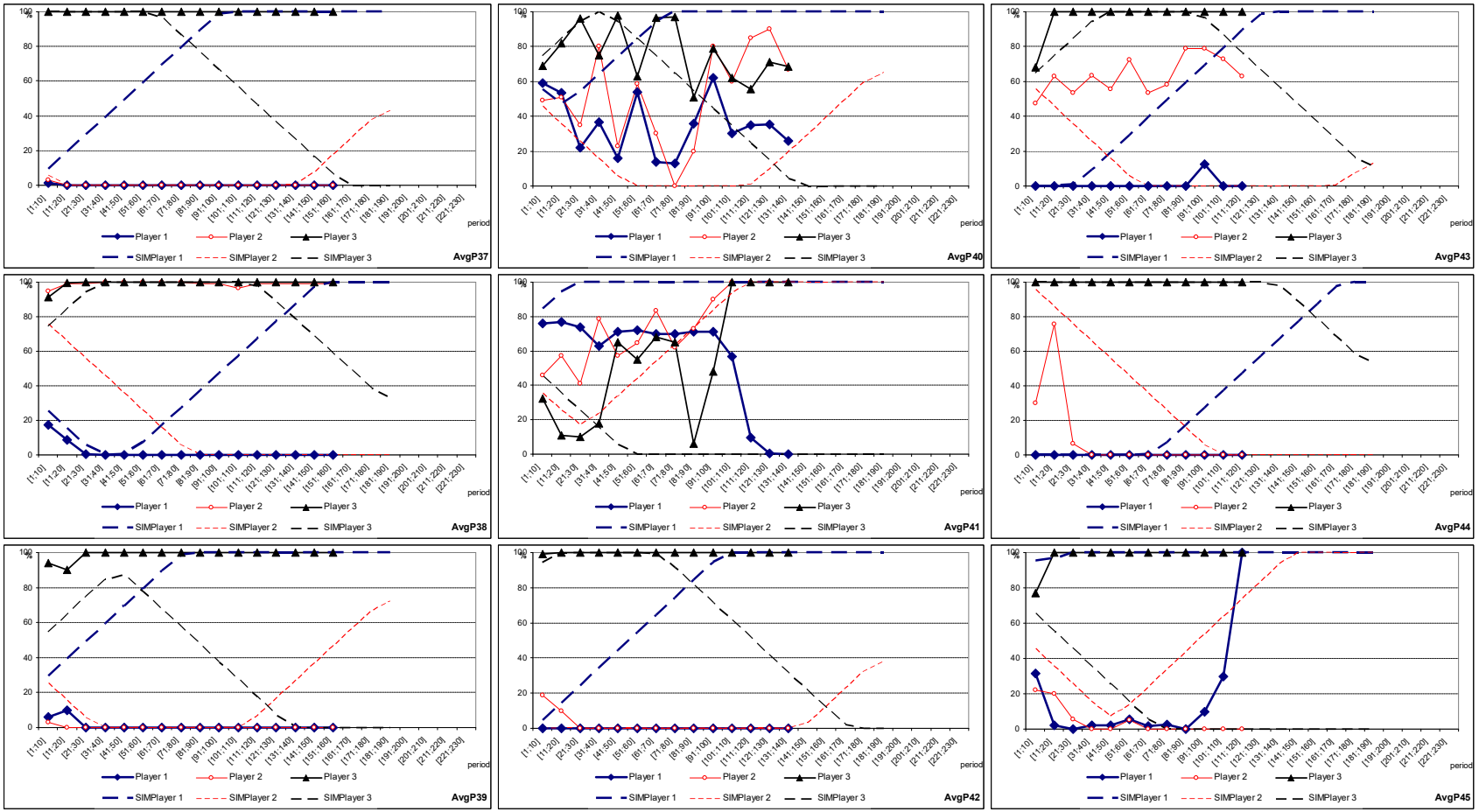


Figure A3.3 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Average Pay treatment of the long-run partners' experiment: average behavior strategies, probability of playing R over 10 periods

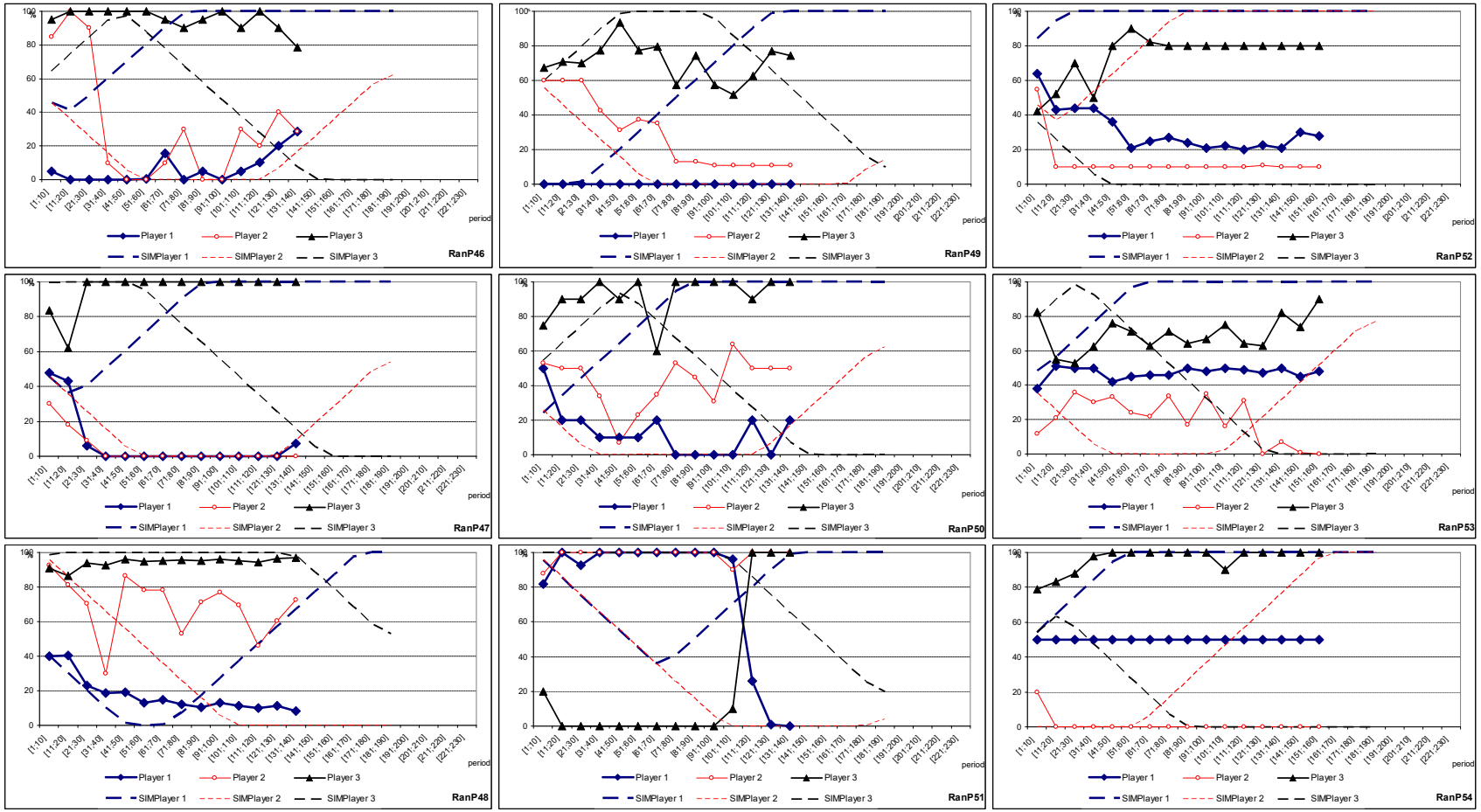


Figure A3.4 Observed trajectories (solid lines) and impulse response simulation (broken lines) in Random Pay treatment of the long-run partners' experiment: average behavior strategies, probability of playing R over 10 periods