Learning in Contests with Payoff Risk and Foregone Payoff Information*

Aidas Masiliūnas†

April 16, 2020

Abstract

We test whether deviations from Nash equilibrium in rent-seeking contests can be explained by the slow convergence of payoff-based learning. We identify and eliminate two sources of noise that slow down learning. The first source of noise is present because each action is evaluated against a different sample of actions of other players. We eliminate it by providing foregone payoff information, which allows all actions to be evaluated against the same sequence of opponent's actions. The second source of noise is present because of payoff risk, which reduces the correlation between expected and realized payoffs. We manipulate payoff risk using a 2x2 design: payoffs from contest investments are either risky (as in standard contests) or safe (as in proportional contests), and payoffs from the part of endowment not invested in the contest are also either safe (as in standard contests) or risky. We find that Nash equilibrium rates go up to 100% when payoff risk is not present and foregone payoff information is available, but are at most 20% in all other cases. This result can be explained by payoff-based learning but not by other theories that might interact with payoff risk (non-monetary utility of winning, risk-seeking preferences, spitefulness, probability weighting, QRE). We propose a hybrid learning model that combines reinforcement and belief learning with preferences, and show that it fits data well, mostly because of reinforcement learning. Additional support for learning comes from the persistence of the Nash equilibrium following the removal of foregone payoff information.

Keywords: experiment, contests, reinforcement learning, foregone payoffs, payoff risk, Nash equilibrium

JEL classification: C72 C91 D71 D81

[*] I would like to thank participants at seminars and conferences in Maastricht (BEElab meeting), Vienna (ESA European meeting 2017), Gothenburg (12th Nordic Conference on Behavioral and Experimental Economics), Marseille (Eco-lunch seminar), Bath (The Micro and Macro Foundations of Conflict), Brno (Young Economists' Meeting), Nice (ASFEE 2018), Vilnius (Baltic Economic Conference), Berlin (ESA World Meeting), Singapore (Game Theory and Social Dynamics Workshop at SUTD). Financial support from Aix-Marseille School of Economics and GSBE at Maastricht University is gratefully acknowledged.

[†] Global Asia Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. E-mail: aidas.masiliunas@gmail.com

1 Introduction

The discrepancy between choices in experiments and Nash equilibrium prediction is often attributed to non-standard preferences or decision error (Goeree and Holt, 2001, DellaVigna, 2009). But in repeated games, the strongest support for Nash equilibrium comes from it being the long-run outcome of a learning process (Mailath, 1998, Perea, 2007), which may fail to converge in relatively short experiments. Slow convergence is especially problematic for payoff-based learning models (Hopkins, 2002) if little can be learnt from the realized payoff information due to the payoff noise (Thrun and Schwartz, 1993, Hasselt, 2010). Payoff noise can originate either from a stochastic payoff function or from the inter-temporal variability of opponent's choices. We study a rent-seeking (Tullock) contest in which both sources of noise are present and identify how they affect the adaptation process. We eliminate these sources of noise by removing the probabilistic prize allocation rule and providing foregone payoff information (FPI) and test whether these manipulations increase the commonly observed low explanatory power of Nash equilibrium.[1]

The first source of noise is the stochastic payoff function when decisions are made under risk. Game theory offers little explanation for how the payoff risk affects the explanatory power of Nash equilibrium in repeated games,[2] perhaps because decisions in most games are made only under strategic uncertainty and not risk. Payoff risk has been studied in individual choice experiments, both one-shot (e.g. Kahneman and Tversky, 1979) and repeated (often using the "decisions-from-experience" paradigm[3]). This literature finds that payoff risk reduces the rates of payoff maximization (Myers and Sadler, 1960, Erev and Barron, 2005) and the expected payoff-maximizing action is rarely chosen when payoff risk is present.[4] Given these findings, the low explanatory power of a Nash equilibrium in contests should not be surprising, as contests are typically run with a larger strategy space, lower optimization incentives, a smaller number of repetitions and variability in opponent's behaviour.[5]

In standard contests (our SR treatment), payoff risk is present because of a probabilistic prize allocation rule. We eliminate it by paying the expected value of the lottery (SS treatment). However, convergence to Nash equilibrium in this treatment could occur for reasons other than payoff-based learning, as the removal of payoff risk might eliminate the effect of risk preferences, probability weighting or non-monetary utility of winning. To better understand the mechanism through which payoff risk operates, we designed two additional treatments with different payoff risk manipulations. In the reverse contest (RS treatment), payoff risk is reversed by making contest investments safe, but the amount not invested in the

[1] See Sheremeta (2013) for an overview of overbidding in contests and potential explanations.

[2] It has been widely studied how risk alters Nash equilibrium predictions by entering the utility function. Instead, we are interested in how risk affects the convergence process in repeated games. We will compare this channel to the risk preference channel in section 4. Two notable exceptions that studied the effect of risk on convergence are Bereby-Meyer and Roth (2006) and Shafran (2012).

[3] "Decisions-from-experience" paradigm is used mainly in psychology experiments, in which players choose between several options that generate payoffs from an unknown distribution. Typically no information about the nature of the payoff distribution is provided, to test how choices are made only from experience, in contrast to "decisions-from-description" (for an overview, see Erev and Haruvy, 2016).

[4] For example, in one task run by Grosskopf et al. (2006) payoffs are drawn from one of the two distributions ($\mathcal{N}(11, 1)$ and $\mathcal{N}(10, 3)$) and the expected payoff-maximizing option is chosen only about half of the time.

[5] For example, in Grosskopf et al. (2006) there are two options, 200 trials and the expected payoff difference of 9%. All contest experiments have much a much larger strategy space, none are repeated more than 60 times (Fallucchi et al., 2013) and 7% of the expected payoff is lost from overbidding the Nash equilibrium level by 100% (median overbidding rate in contest experiments reviewed by Sheremeta, 2013, is 72%).

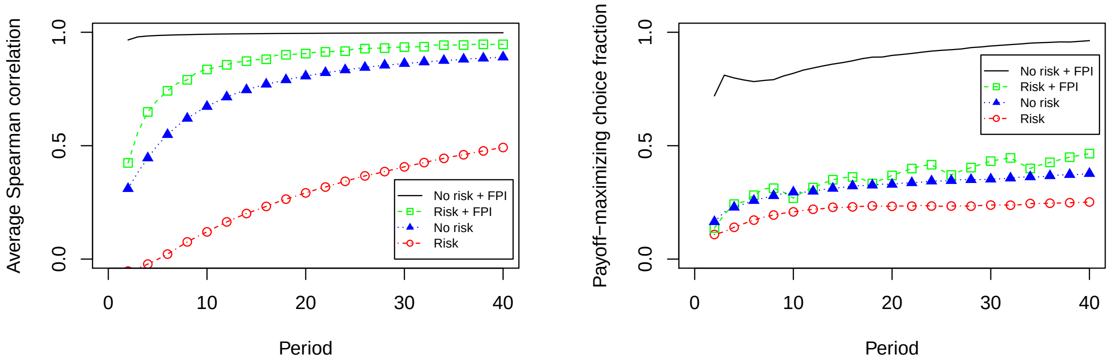
contest risky. In RR treatment, both types of investment are risky. Risk preference hypothesis predicts Nash equilibrium play only in SS. Non-monetary utility of winning and probability weighting predict Nash equilibrium play in SS and in RR. Spitefulness and inequality aversion predict similar rates of deviations from the Nash equilibrium in all four treatments. This variability in predictions allows us to evaluate how well each theory can explain the data from all treatments.

Even when the payoff risk is removed, the convergence of payoff-based learning is predicted to be slow because of the inter-temporal variability of opponents' choices. Payoff-based learning converges through the iterated elimination of dominated strategies (Beggs, 2005), but the process is slow if opponents are changing their actions over time and thus the payoffs of dominated actions can exceed the payoffs of actions that dominate, even in the absence of payoff risk. We eliminate the effect of this variability by providing FPI, i.e. revealing what payoffs would have been generated by unchosen actions. This manipulation ensures that the variability of opponent's actions exerts a uniform effect on the performance of all own actions.

As an illustration, we ran a simulation to see how long players would need to play the contest to discover which actions produce the highest payoffs, if only learning from experience was possible. To simplify, assume that contest is played as a multi-armed bandit task, with each of the nine arms representing a contest investment level (0-8). When an arm is played, opponent's action is drawn from a stationary distribution (equal to the distribution of choices in our baseline experiment) and a profit is generated using a contest success function (with or without payoff risk).[6] Each round, one arm is chosen at random (uniformly) and the payoff is observed. The procedure is performed under one of the four conditions, corresponding to our experimental design: payoff risk is either present or not and FPI is either observed or not. For each condition, we calculate the average observed payoff that each arm generated over all previous rounds.[7] We then measure how well the average observed payoff reflects the expected payoff of that arm. First, we use Spearman's correlation coefficient to measure how well the ranking of arms in terms of average observed payoffs recovers the ranking based on expected payoffs. Panel (a) of figure 1 shows the average correlation from 10 000 simulations. In the standard contest, the correlation is very low and increases only to 0.5 after 40 rounds of experience. Correlation is increased if payoff risk is removed or if FPI is provided, and both manipulations combined produce a nearly perfect positive correlation after just one round. Second, we calculate how often the arm that maximizes average observed payoffs coincides with the arm that maximizes objective expected payoffs. This measure can be interpreted as the likelihood that a Bayesian player will correctly identify the expected payoff-maximizing action using all previous payoff realizations. Panel (b) in figure 1 shows that expected payoff maximization fails at least 50% of the time if either payoff risk is present or forgone payoffs are not observed, even after 40 rounds. In contrast, the maximization rate approaches 100% with no payoff risk and with FPI. In terms of both correlation and payoff maximization, feedback from one round with FPI and no payoff risk improves the decision quality more than 40 rounds of experience in either of the other three conditions. We ran additional simulations to see how soon the payoff maximization rate reaches 73%, the rate achieved in just one round with no payoff risk and with FPI. It takes about 500 rounds with payoff risk and FPI, 1500 rounds without payoff risk and without FPI and 15 000 rounds with payoff risk and with FPI.

[6] A game with such structure has been implemented by Cox (2017), finding no difference at the aggregate level between this "robot" and standard "human" treatment.

[7] If no payoff is observed, we use the average expected payoff from the task, equal to 8.02.



(a) Average correlation between expected payoffs and average observed payoffs over prior rounds

(b) Expected payoff maximization, if choice is based on highest observed average payoff.

Figure 1: Simulated recovery of expected payoffs from observed payoffs in a bandit task

Overall, the simulation shows that realized payoffs in contests are a noisy measure of expected payoffs,[8] and players who wanted to maximize expected payoffs would fail to do so if decisions were made only from experience. The removal of payoff risk and provision of FPI reduce the sampling error by implementing "full reinforcement", that is reinforcing all actions all the time, while under other conditions reinforcement is partial. In individual choice tasks, it has been shown that full reinforcement improves expected payoff maximization due to faster payoff-based learning (Erev and Haruvy, 2016). We show that in contests too, payoff-based learning predicts high rates of Nash equilibrium play only under full reinforcement.

Experimental results confirm this prediction. In the treatment with no payoff risk, Nash equilibrium rates reach 100% when FPI is introduced, compared to at most 20% in the other three cases. Reinforcement learning also correctly predicts that once Nash equilibrium is reached, it persists even when FPI is removed. In contrast, we find little difference between treatments when FPI is not available. To find which theory can account for these results, we estimate a learning model that combines reinforcement and belief learning with probability weighting, non-monetary utility of winning, risk and social preferences. The full model fits better than the generalized experience-weighted attraction (EWA), belief or reinforcement learning models because reinforcement can explain high equilibrium rates in the treatment with no payoff risk following the introduction of FPI, while non-monetary utility of winning and probability weighting can explain treatment differences between treatments in which payoff risk is present.

We make several contributions to the literature that studies the origins of overbidding in contests. Explanations for why behavior in lab and field experiments differs from theoretical predictions are often divided into three categories: non-standard preferences, non-standard beliefs or non-standard decision making process (we use the classification and language from DellaVigna, 2009). If non-standard preferences[9] are behind deviations from theoretical predic-

[8] In the simulation, expected payoffs can be calculated because actions of opponents are drawn from a stationary distribution. In strategic games, expected payoffs could be calculated ex-post. Models such as reinforcement learning converge through iterated elimination of dominated strategies, therefore their speed depends on how accurately the realized payoff differences reflect differences in terms of expected payoffs.

[9] We say that people have "non-standard preferences" if their utility function differs from the expected payoff function, for example because of risk, social preferences or non-monetary utility of winning.

tions in contests, then such deviations should disappear when those preferences are eliminated by design. Some studies followed this approach, paying the expected value of the lottery to remove the effect of risk preferences or matching participants with robot opponents to switch off social preferences (Chowdhury et al., 2014, Cox, 2017, Masiliūnas et al., 2014). It is generally found that such manipulations do not reduce overbidding. Non-standard beliefs about the behavior of other participants could be corrected by playing the contest sequentially (Fonseca, 2009) or by informing participants about the action that will be played by a robot opponent (Masiliūnas et al., 2014). It has been found that such treatments produce similar behavioral patterns as the standard contest game. If non-standard decision making is behind the deviations, how they can be reduced will depend on the nature of the decision making process. If participants attempt but fail to optimize when making decisions from the description of the game, decision quality could be improved by providing payoff calculators or explaining the game in a more understandable way (Chowdhury et al., 2020). Our paper is the first to focus on the process of making decisions from experience. If people attempt to learn from their own past experience, but fail because feedback is not sufficiently informative, then the quality of decisions could be improved by providing higher quality feedback. We find that it does, but only in the absence of payoff risk, as predicted by a payoff-based learning model. We also find that payoff-based learning can organize experimental data better than the explanations based on non-standard preferences. We use the data from all our treatments to estimate a learning model that combines non-standard preferences with learning from experience. This novel method of jointly modeling bounded rationality and non-standard preferences could be useful to disentangle competing mechanisms is future experiments with different games.

2 Literature

2.1 Deviations from Nash equilibrium in contests

In contest experiments (see a survey by Sheremeta, 2013), average choices almost always exceed the Nash equilibrium prediction ("over-investments", "over-expenditure", "overspending" or "overbidding") and the distribution of choices is widely spread ("overspreading", Chowdhury et al., 2014), therefore Nash equilibrium fails to organize experimental data (Sheremeta and Zhang, 2010, Masiliūnas et al., 2014). The gap between behavior and theoretical predictions is usually explained by assuming either non-standard preferences or bounded rationality.

First, players may have a preference or aversion to risk. Failure of the risk neutrality assumption can explain the difference between theoretical predictions and experimental data in related games, such as auctions (Goeree et al., 2002, Cox et al., 1988). Contest investments have elements of both gambling and insurance, therefore it is not possible to identify the effect of risk preferences without assuming a specific utility function, usually with constant absolute risk aversion (CARA) or constant relative risk aversion (CRRA). Hillman and Katz (1984) show that risk aversion reduces investments if all participants use a logarithmic utility function (which exhibits CARA). Skaperdas and Gan (1995) assume exponential utility functions (which also exhibits CARA) and find that the more risk averse agent invests less than the opponent as long as both agents are sufficiently similar in terms of their risk preferences. Cornes and Hartley (2003) show that under exponential utility total investments are always lower when all agents are either risk averse or risk neutral, compared to full risk neutrality. Jindapon and Yang (2017) show that with a generalized CARA utility function both risk averse and risk seeking players invest below the risk-neutral Nash equilibrium prediction in symmetric simultaneous two-player contests. Jindapon and Whaley (2015) use both exponential (CARA) and CRRA utility functions that exhibit risk-seeking preferences and show that in two-player contests with identical preferences imprudence increases over-investments, irrespective of the risk-aversion coefficient. Another strand of literature tests whether heterogeneity in choices can be explained by heterogeneity in risk preferences, elicited in experiments. It is typically found that risk averse participants on average invest less than risk seeking participants (Millner and Pratt, 1991, Herrmann and Orzen, 2008, Sheremeta and Zhang, 2010, Sheremeta, 2011, Price and Sheremeta, 2015).

The second preference-based hypothesis assumes other-regarding preferences. In theory, over-investments could be explained by spiteful preferences (Riechmann, 2007), aversion to disadvantageous inequality, or a preference for advantageous inequality (Herrmann and Orzen, 2008, Fonseca, 2009). Some studies elicit spitefulness or inequality aversion and correlate it with behavior in contests. Savikhin and Sheremeta (2013) find that participants who contribute more in a public goods game invest less in the contest. Herrmann and Orzen (2008) measure pro-sociality using a prisoner's dilemma and find the opposite result: pro-sociality is correlated with higher contest investments. An alternative to measuring social preferences is to eliminate their effect by removing the payoff consequences to the opponent. Herrmann and Orzen (2008) use a strategy method and find higher investments when playing against other participants compared to playing against no competitor. Masiliūnas et al. (2014) and Cox (2017) match players to computers that are programmed to play choices made by participants in the standard contest experiment. Both studies find no significant difference between computer and human treatments.

The third theory assumes that winning provides non-monetary utility, in addition to the monetary value of the prize (Schmitt et al., 2004). Sheremeta (2010) measures non-monetary utility of winning by asking players to compete for a prize of zero value. It is found that investments in zero-value contests are correlated with investments in standard contests (Price and Sheremeta, 2011, Price and Sheremeta, 2015, Brookins and Ryvkin, 2014, Sheremeta et al., 2018, Mago et al., 2016, Cox, 2017). Such correlation does not necessarily imply causality, as both could arise because of confusion, experimenter demand effect, mistrusting the experimenter or habit (see Sheremeta, 2010, and Sheremeta, 2013, for a discussion).

Other theories explain over-investments through bounded rationality. It is well known that biases occur when choosing between risky gambles. In particular, it is often found that objective probabilities are weighted using an inverted S-shaped function (Tversky and Kahneman, 1992, Gonzalez and Wu, 1999). Such probability weighting could explain over-investments in contests with many players, but it predicts under-investments in two-player contests (Baharad and Nitzan, 2008) and fails to explain the data (Parco et al., 2005). Instead, over-investments could be rationalized by S-shaped probability weighting, which can originate from experience-based decision making (Hertwig et al., 2004).

Perhaps the most commonly used model of bounded rationality is the quantal response equilibrium (QRE), which assumes that players are not perfectly sensitive to expected payoff differences (McKelvey and Palfrey, 1995). QRE can explain over-investments because decreased sensitivity increases the variance of the choice distribution. It has been found that QRE can explain some choice patterns in contests (Sheremeta, 2011, Lim et al., 2014, Brookins and Ryvkin, 2014). Other evidence for bounded rationality comes from increased explanatory power of Nash equilibrium when game difficulty is reduced by removing payoff risk and strategic uncertainty (Masiliūnas et al., 2014) or by using an explicit ticket-based lottery implementation (Chowdhury et al., 2020).

Theories based on non-standard preferences are typically evaluated by eliciting the preference and testing if it is correlated with contest investments. This approach has several shortcomings, one of them being the risk of omitted variable bias. Sheremeta (2016) reduces the bias by measuring multiple personal characteristics in a single experiment, and testing which ones can best explain choices in a one-shot contest. When tested one-by-one, many variables are significant: investments are higher for players who state higher beliefs, make more mistakes in the quiz, are less loss-averse, invest more in the zero-value contest, are more competitive and make more mistakes in a cognitive reflection test. Other characteristics, such as risk aversion, cognitive abilities, and violation of expected utility theory, are not significant. When all preferences are combined in a single model, only the performance in the cognitive reflection test remains highly significant, beliefs and mistakes in the understanding quiz are marginally significant, and all other variables are not.

2.2 Payoff risk and foregone payoff information

Contests with no payoff risk (equivalent to the SS treatment) have been studied by Fallucchi et al. (2013), Chowdhury et al. (2014) and Masiliūnas et al. (2014).[10] All three studies find that the removal of payoff risk does not reduce the rate of over-investments, although the magnitude of over-investments and under-investments tends to be reduced. In contrast, the removal of payoff risk reduces over-investments and increases the explanatory power of Nash equilibrium if the optimization premium is high (Chowdhury et al., 2014), if information about individual choices of other participants is withheld (Fallucchi et al., 2013) or if the opponent is required to play the same action for some time (Masiliūnas et al., 2014).

We are not aware of literature that manipulates the availability of FPI in a strategic game,[11] but some designs have been used with a similar effect. Of particular relevance are the studies that vary payoff information and feedback. The first strand of literature manipulates the format of information given to the participants, hypothesizing that if bounded rationality is behind deviations from theoretical predictions, then an improved understanding of the incentive structure should reduce the magnitude of such deviations. Support for this hypothesis has been found in gift exchange games (Charness et al., 2004) and Cournot oligopoly (Bosch-Domènech and Vriend, 2003). Our design is different because we change the feedback participants receive after each round, in contrast to changing the information that participants have prior to making their decisions. We thus hypothesize that this manipulation will affect decisions from experience rather than decisions from description. The second strand of literature targets decisions from experience, just as we do, by varying the feedback participants receive. For example, it has been studied how the adaptation behavior changes when participants learn the actions and payoffs of their opponents (in public goods game, Nax et al., 2016, market entry games, Duffy and Hopkins, 2005). Similar feedback conditions have been studied in first price auctions, revealing the second highest bid to the winner or the highest bid to the loser (Filiz-Ozbay and Ozbay, 2007; Engelbrecht-Wiggans

[10] Other studies look at proportional contests with a slightly different design: in Sheremeta et al. (2018) the cost of effort is quadratic and investments are subject to additional noise, in Shupp et al. (2013) the strategy space is censored, which seems to be the reason why average investments are below the Nash equilibrium.

[11] To the best of our knowledge, the only study that manipulated FPI in a strategic game is Grosskopf et al. (2007), who do so in the "acquiring a company" task. The game is strategic only in control treatments, in which the role of the seller is assumed by human participants, but their decisions are trivial.

and Katok, 2009).[12] Feedback about opponent's bid allows the calculation of foregone payoffs in all such deterministic games. Our experiment is different because information about the actions of other participants is available in all treatments, thus foregone payoffs can always be calculated. However, doing so would be difficult because of the computational complexity, which is why we went a step further and provided FPI explicitly.

FPI has been studied in individual choice tasks, typically using the "decisions-from-experience" paradigm. Rakow et al. (2015) find that the alternative with a higher expected value is chosen more often when FPI is available. Grosskopf et al. (2007) and Fudenberg and Peysakhovich (2016) find that sub-optimal overbidding in "acquiring a company" game and "additive lemons problem" does not reduce when FPI is provided. Grosskopf et al. (2006) show that FPI increases expected payoff maximization if payoffs from both actions are positively correlated, but reduces it in other cases, a result attributed to over-exploration and increased focus on large positive payoffs generated by the more risky option ("big eyes effect"). Otto and Love (2010) find that in a dynamic game FPI reduces the choice of an option that maximizes long-run payoffs, in favour of an option that provides a high immediate payoff. This result can be explained by reinforcement learning, because FPI increases the attractiveness of the option with high immediate payoffs and the strategy space remains under-explored. Yechiam and Busemeyer (2006) investigate a task in which the more risky action usually generates a higher payoff, but has a lower expected payoff because of a small probability of a large loss. The risky option is chosen more often when FPI is available, and the treatment difference is larger when the probability to receive the large negative reward is small (1/200 compared to 1/20). These results can be well explained by reinforcement learning. Overall, this literature shows that the effect of FPI on the payoff maximization rates is task-specific, and can be predicted by reinforcement learning. Our study contributes to this literature by testing the reinforcement learning predictions about the effect of FPI on Nash equilibrium play rates in games with varying sources of payoff risk.

3 Experimental design

In the standard rent-seeking contest, the part of endowment invested in the contest is risky, while the part not invested is directly converted into earnings, and therefore not risky. We designed three additional contest variations in which players also choose how to divide the endowment, but the riskiness of contest and non-contest investments is manipulated using a 2x2 between-subject design. Treatment differences are displayed in figure 2.

- **Treatment SR** (*safe* non-contest investments, *risky* contest investments, or "standard") is equivalent to the standard contest game.
- **Treatment SS** (*safe* non-contest investments, *safe* contest investments) is equivalent to a proportional contest (Masiliūnas et al., 2014, Fallucchi et al., 2013). The payoff risk from contest investments is removed by paying the expected value of the lottery.
- **Treatment RS** (*risky* non-contest investments, *safe* contest investments, or "reverse") makes income from non-contest investments risky and income from contest investments

[12] These studies are different from our design not only because of a different game, but also because they do not study repeated strategic games: Filiz-Ozbay and Ozbay (2007) use a one-shot design while participants in Engelbrecht-Wiggans and Katok (2009) compete against robots.

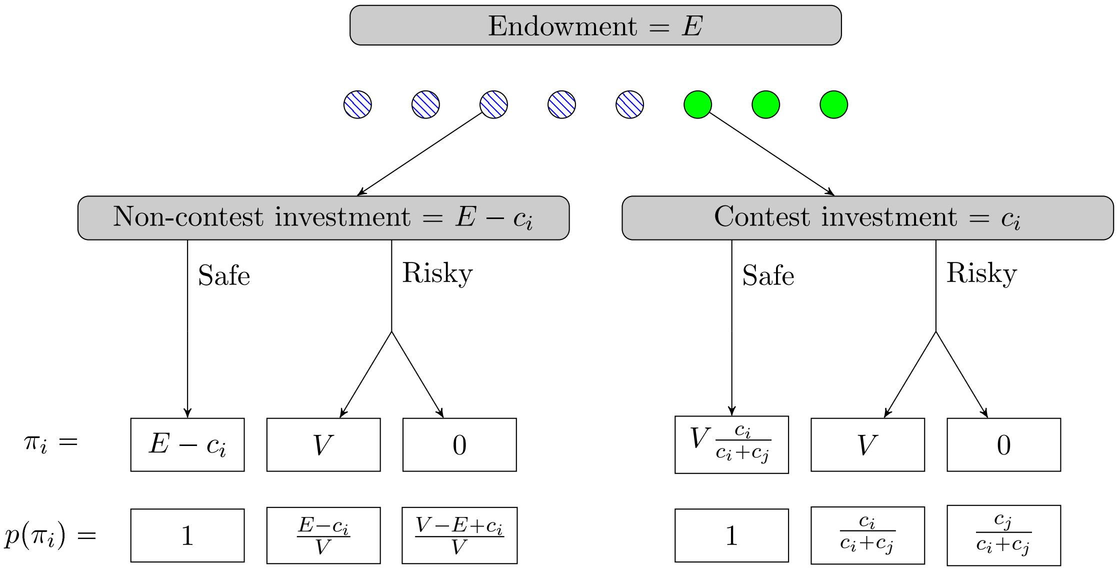


Figure 2: Sources of payoff risk in four contest treatments.

safe, reversing the payoff risk compared to the SR treatment. In RS, contest investments generate a deterministic payoff, just as in SS, but non-contest investments determine the probability to win a prize. The prize value is the same as in SR, but the odds to win are proportional to non-contest investments. We set the prize and endowment to 8 points, therefore each unit of non-contest investment increased the probability to receive the prize by 12.5 percentage points.

- **Treatment RR** (*risky* non-contest investments, *risky* contest investments) combines two lotteries with identical prize values: in one the probability to win is determined by contest investments, in the other by non-contest investments.

Besides the payoff risk, we also did a within-subject manipulation of foregone payoff information (FPI). All treatments started with two non-incentivized rounds, used to familiarize participants with the software, followed by 10 incentivized rounds in which FPI was not available, 20 rounds with FPI and 10 more rounds without FPI. The first 10 rounds provide a baseline comparison of treatments, with no additional information. The block with FPI had 20 rounds because it is the focus of the paper. If the payoff-based learning hypothesis is correct, a treatment difference should appear at this stage. The third block tests whether behavior observed in an environment with FPI persists when FPI is removed. If players consider the entire history of outcomes (as in reinforcement learning), behavior in the third block should be similar to the second block, and treatment differences would persist. If, on the other hand, behavior is affected only by immediate feedback (as in regret minimization, or directional learning), behavior and treatment order in the third block would resemble the first block. At the end of the experiment, we collected demographic data and asked what action participants would recommend to a friend who would hypothetically take part in this experiment (adapted from Grosskopf et al. 2007). Responses to this question provide further insight into what players learn from the game, without the potential confounds of reputation building or exploration-exploitation tradeoff.

We varied the availability of FPI over time, instead of using a between-subject design, because we were interested in how the appearance of FPI alters the adaptions process on an individual level, and whether information has a similar effect on all participants. Information about individual-level adaption is especially useful to construct an accurate learning model that can explain the consequences of providing additional feedback. Another benefit of the chosen design is the relevance to real-life problems, since policymakers are interested in policies that alter undesirable behaviours and have a long-term effect. An important element of this challenge is that habits are already in place, and a policy change needs to overcome them. The within-subject design captures the problem by first inducing behavior typically observed in the previous literature, and then testing if it can be corrected by providing appropriate information (for similar designs in the literature, see Jiao and Nax, 2018, or Masiliūnas, 2017). On the other hand, introducing FPI over time creates some challenges. First, the roll-out of FPI could induce an experimenter demand effect, as participants might feel pressured to adjust their behavior. When analyzing the results, we take this into account, focusing on the comparison between the four treatments, which should be similarly affected by the demand effect. Second, since FPI is changed over time, it is difficult to conclude whether behavioral change is caused by FPI or whether it would have occurred even without any additional interventions. We tackle this problem in the analysis part as well, comparing the trends in our experiments to similar previous experiments that did not change anything over time. Overall, we are primarily interested in the comparison between the four treatments (either with or without FPI), and less so in the comparison before and after the introduction of FPI, thus the challenges associated with the within-subject design can be deal with.

A graphical interface was used to unify the presentation of all treatments. The computer screen seen by participants is reproduced in figures 18-23 (Appendix H).[13] The decision screen (figure 18) was identical in each round and each treatment, but the presentation of feedback varied. In each treatment, income from the contest and non-contest investments was depicted as a box containing eight squares, the highest possible payoff from each investment type. In SS, one square was added into the non-contest box for each unit not invested into the contest, while the contest box was divided proportionally to contest investments. In SR, the screen looked similar, but the share of the contest box represented the probability to win. After players had observed their probability to win, a lottery was performed by placing a marker at a random location of the contest box. If the marker stopped in the player's area, the prize was won and eight squares were added to round income. In RS, squares in the non-contest box represented the probability to win, and the lottery was performed by placing a marker at a random location of the non-contest box. If the marker stopped in the player's area, the prize was won and eight squares were added to the income. In RR, two lotteries were performed, one for contest and one for non-contest investments. Two markers stopped at random locations of each box, and a player received zero, one or two prizes.

In addition to the payoff information, players received feedback about the action of the other participant, the received share of the contest prize in percentage and in points (in SS and RS) or the probability to win the contest prize (in SR and RR). When FPI was available, players were additionally informed about the probability to win and lottery outcomes for the actions that they did not choose. The same randomly generated number was used to determine lottery outcomes for all actions. To view FPI, participants had to click on a box

[13] A video illustrating the graphical interface seen by participants can be viewed at `https://www.dropbox.com/s/m2quh6ku4lsad7y/learning_in_contests_video.mp4?raw=1`

that was covering the information. Information revelation was recorded, allowing us to know what information players observed.

In each round, players were randomly rematched within a 6-person matching group. The prize and the endowment were set to 8 points and players could invest integer amounts between 0 and 8 points.[14] These parameters ensure that the Nash equilibrium is an integer amount and the strategy space is sufficiently small to display the FPI on the computer screen. Four rounds were randomly chosen for payment, one from each block of 10 rounds. Players received instructions for each of the three parts only at the start of that part, so that decisions in the first part would not be influenced by information about the subsequent introduction of FPI.

A total of 144 participants took part in the experiment, 36 in each of the four treatments. The average duration of the experiment was 80 minutes, and the average payment was €16.50. Experiments were programmed using z-Tree (Fischbacher, 2007) and run at the BEElab of Maastricht University. Subjects were recruited using ORSEE (Greiner, 2015).

4 Theoretical predictions

Each player i divides the endowment E between contest investment (c_i) and non-contest investment ($E - c_i$). We will refer to c_i as the choice variable. Depending on the treatment, investments determine either the share of the prize or the probability to receive the prize. Denote the value of the contest and non-contest prize by V, the probability/share of the contest prize by p_i^c and the probability/share of the non-contest prize by p_i^{nc} such that:

$$p_i^c(c_i, c_j) = \begin{cases} \frac{c_i}{c_i + c_j} & \text{if } c_i + c_j > 0 \\ 0.5 & \text{otherwise} \end{cases}$$

$$p_i^{nc}(c_i) = \frac{E - c_i}{V}$$

Table 1 shows the prospects that players face in each treatment. The expected payoffs are identical in all treatments:

$$E[\pi_i(c_i, c_j)] = E - c_i + V \frac{c_i}{c_i + c_j}$$

Table 1: Treatments and corresponding payoff functions. The endowment equals E, the prize equals V, contest investment of player i is c_i and $\{V, p\}$ denotes a gamble that pays V with probability p and 0 with probability $(1 - p)$.

		Contest investment	
		Safe	Risky
Non-contest investment	Safe	SS $\pi_i = (E - c_i) + V\frac{c_i}{c_i + c_j}$	SR $\pi_i = (E - c_i) + \{V, \frac{c_i}{c_i + c_j}\}$
	Risky	RS $\pi_i = \{V, \frac{E - c_i}{V}\} + V\frac{c_i}{c_i + c_j}$	RR $\pi_i = \{V, \frac{E - c_i}{V}\} + \{V, \frac{c_i}{c_i + c_j}\}$

[14] In the experiment, all earnings were denominated in "points" with an exchange rate of 1 point = €0.45.

A unique stage-game Nash equilibrium in all four treatments is $(c_i^*, c_j^*) = (V/4, V/4)$, reducing to $(2,2)$ when we set $E = V = 8$. We use the term "risk-neutral Nash equilibrium" (RNNE) for the Nash equilibrium calculated under the assumption of standard preferences, to avoid confusion with the Nash equilibria calculated with non-standard preferences. RNNE predictions are invariant to changes in payoff risk (because expected payoffs are held constant) and FPI (because the payoff function is known in all treatments), therefore RNNE predicts no difference between treatments and between rounds with and without FPI.

Contest investments above RNNE (termed "over-investments") are strictly dominated by the RNNE action. With the discrete strategy space $[0, 1, \ldots, 8]$, only the RNNE action profile survives the iterated elimination of dominated strategies.[15]

Behavior in contest experiments is systematically different from RNNE predictions (Sheremeta, 2013), and this low explanatory power could be attributed to restrictive assumptions about rationality or preferences. The remainder of this section will specify how the theoretical predictions change when RNNE assumptions are relaxed, permitting bounded rationality or non-standard utility functions. Bounded rationality is modeled using intermediate run predictions of a payoff-based learning model, Quantal Response Equilibrium and Impulse Balance Equilibrium. Alternatively, we keep the concept of Nash equilibrium, but enrich the expected utility function with risk and social preferences, probability weighting and non-monetary utility of winning.

4.1 Payoff-based learning

The calculation of RNNE from the description of the game is cognitively demanding: participants would need to calculate the expected payoffs at each action profile and either eliminate actions that cannot be supported by any belief hierarchy or find a best-response to each possible action of the opponent. An alternative justification for the Nash equilibrium arises from it being the long-run outcome of learning dynamics, such as of belief learning (Lehrer and Kalai, 1993) or replicator dynamics (Weibull, 1997). Belief learning is less cognitively demanding than iterated elimination of dominated strategies, as it requires expected payoffs to be calculated not for all action profiles but only once for each action. Still, best-responding is difficult and in experiments the theoretical best-response is rarely chosen even when opponent's action is known (Fonseca, 2009, Masiliūnas et al., 2014). The most likely path to RNNE under bounded rationality is through payoff-based learning. The main challenge lies not with the cognitive abilities of the participants, but with the quantity and quality of observed feedback.

"Payoff-based" or "completely uncoupled" learning depends only on player's own past payoffs (Foster and Young, 2006). It has been shown that in some games some models in this class converge to RNNE (aspiration learning, or win-stay, lose-shift, Cho and Matsui, 2005, regret testing, Foster and Young, 2006, trial and error, Young, 2009, probe and adjust, Huttegger, 2013, reinforcement learning, Beggs, 2005). We use the most popular model in this class, reinforcement learning (Erev and Roth, 1998), which has been successfully used to explain deviations from expected payoff-maximization in individual choice tasks with FPI (Grosskopf et al., 2006, Otto and Love, 2010, Yechiam and Busemeyer, 2006, summarized in section 2). Theoretically, reinforcement learning converges in contests through the iterated elimination of dominated strategies (Beggs, 2005), even when the payoffs are stochastic (Bravo

[15] Iterated elimination of dominated strategies requires three steps: in the first step all actions above 2 are dominated by 2, in the second step 0 is dominated by 2, in the third step 1 is dominated by 2.

and Mertikopoulos, 2017). In practice, learning can be slow and incomplete because of the payoff risk and high behavioral variability.

We changed two aspects of the game that should increase the convergence speed of payoff-based learning. The manipulation of payoff risk makes SS the only treatment in which payoffs are deterministic. However, even in SS the correlation between realized and expected payoffs is low because of the variability of the actions chosen by the opponent. The FPI manipulation solves the problem by enabling a direct comparison of all actions against the same distribution of opponent's actions. Convergence by iterated elimination of dominated strategies should be fast in SS with FPI because dominated strategies always generate low payoffs.

Formally, we obtain the predictions about the treatment difference by simulating the path of play for a reinforcement learning model (Roth and Erev 1995, Erev and Roth 1998). Each action $s^k \in [0, \ldots, 8]$, representing contest investments, has an associated attraction in round t, denoted $A_k(t)$. Initial attractions $A_k(0)$ are set to 0, for all k. The foregone or realized payoff from choosing action s^k in round $t \in [1, \ldots, 40]$ is $\pi_k(t)$. Attractions are reinforced using only realized payoffs in rounds 1-10 and 31-40, and both realized and foregone payoffs in rounds 11-30. The reinforcement function is equal to the difference between payoff and a reference point. We assume that the reference point is equal to the initial endowment (E), which is also the maximin payoff and the average payoff if both players choose all actions with equal probabilities.[16] If the action's payoff is unobserved, the attraction remains unchanged. If it is observed (because it was played or because the foregone payoffs are displayed), the attraction in round t is a weighted average of the attraction in round $(t-1)$ and the received reinforcement:

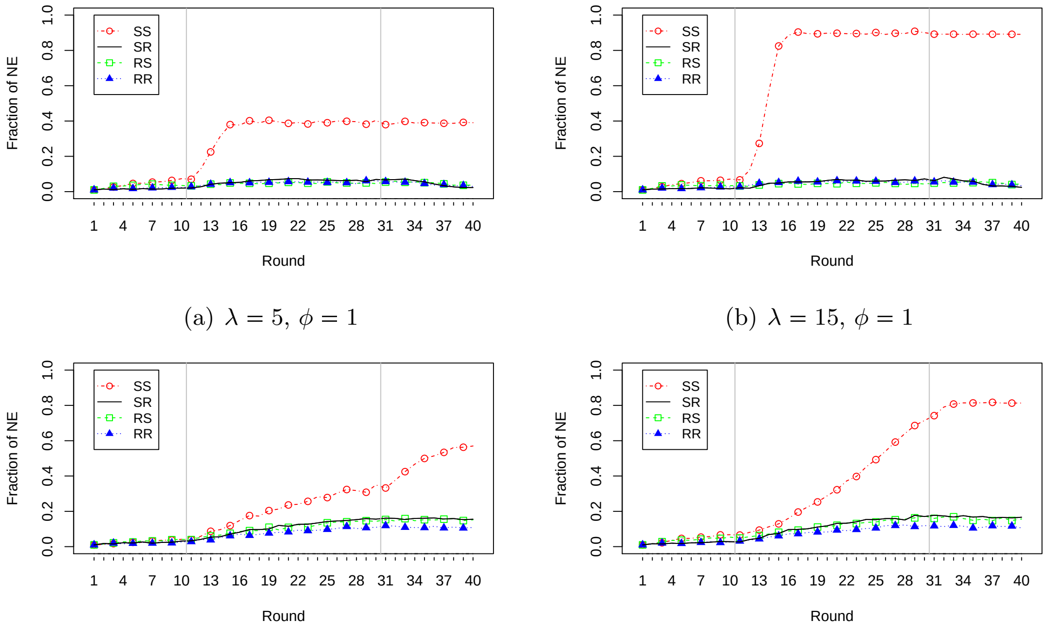
$$A_k(t) = \begin{cases} \frac{\phi A_k(t-1)+\pi_k(t)}{\phi+1} & \text{if } k \text{ is chosen or } t \in [11, \ldots, 30], \\ A_k(t-1) & \text{otherwise.} \end{cases}$$

Attractions are averaged rather than cumulated, to keep the rate of learning constant across rounds. Attractions are mapped into choice probabilities using a logistic choice rule:

$$P_k(t) = \frac{e^{\lambda A_k(t-1)}}{\sum_j e^{\lambda A_j(t-1)}}$$

Reinforcement learning is governed by two parameters: ϕ is the discount factor for older attractions and λ controls sensitivity to differences between attractions. Simulations were run with four combinations of parameter values: the weight placed on previous history was either low ($\phi = 1$) or high ($\phi = 10$), and the sensitivity to differences between attractions was either low ($\lambda = 5$) or high ($\lambda = 15$). Agents were randomly rematched for 40 rounds within a 6-person group, and each simulation was repeated 1000 times. Figure 3 shows the average rate of RNNE play in matched pairs. With all parameter combinations, the simulated difference between treatments is very small in the first 10 rounds, but a difference between SS and the other three treatments appears when FPI is introduced. This treatment difference persists even when FPI is removed. A higher weight placed on recent payoffs (lower ϕ) improves convergence in SS but slows it down in the other three treatments. The difference occurs

[16] We will subtract the reference point E from payoffs in the remainder of the paper, and will use $\pi_k(t)$ to stand for payoffs net of endowment.



(a) $\lambda = 5$, $\phi = 1$

(b) $\lambda = 15$, $\phi = 1$

(c) $\lambda = 5$, $\phi = 10$

(d) $\lambda = 15$, $\phi = 10$

Figure 3: Fraction of pairs choosing RNNE action profiles in reinforcement learning simulations. FPI is observed in rounds 11-30.

because dominated strategies in SS always generate low payoffs, therefore a high weight placed on recently obtained payoffs (low ϕ) and high sensitivity to differences between attractions (high λ) reduces the subsequent probability to choose dominated strategies and leads to fast convergence through the iterated elimination of dominated strategies. In the other three treatments, convergence is slow when recent payoffs receive a high weight (low ϕ) because payoffs from a small number of rounds are noisy due to the probabilistic prize allocation. Payoff noise prevents the elimination of dominated strategies even when λ is high because dominated strategies can generate high average earnings in the short run.

We investigate the robustness of these results by considering a wider set of parameter values. Figure 15 in Appendix G plots the simulated fraction of RNNE pairs in round 10 and round 30. In panel (a), λ is set to an intermediate value of 5, and ϕ is set to values between 0.1 and 20. In panel (b), ϕ is set to an intermediate value of 5, and λ is set to values between 1 and 20. One hundred simulations were run for each parameter combination. Treatment differences are not observed in round 10, but RNNE is more common in SS than in the other three treatments in round 30, for all the considered parameter values.

FPI affects the path of choices because we assume that learning is driven by both realized and foregone payoffs. However, the origins of reinforcement learning lie in behaviorist psychology, which assumes that people respond only to realized payoffs. We interpret reinforcement learning more broadly, as a statistical method to estimate expected payoffs through the aggregation of observed payoff information.[17] This assumption is not uncommon in the

[17] Grosskopf et al. (2006) use the term “fictitious play” to refer to a model that is sensitive to FPI, while “reinforcement learning” refers to a model that does not depend on foregone payoffs. Their formulation of the “fictitious play” model is almost identical to the one presented in this section (only with an added separation

literature; for example, experience-weighted attraction model (Camerer and Ho, 1999) includes a parameter measuring sensitivity to FPI ("law of simulated effect"), and sensitivity to FPI is used in reinforcement learning models to explain choices in experiments with full feedback (Yechiam and Rakow, 2012, Otto and Love, 2010).

4.2 Quantal Response Equilibrium and Impulse Balance Equilibrium

As an alternative to a simulation of a learning model we consider two stationary models of bounded rationality: impulse balance equilibrium (IBE) and quantal response equilibrium (QRE).

Impulse balance equilibrium (Selten et al., 2005) is based on ex-post rationality. It assumes that players receive an impulse to either decrease their action if it was above the ex-post rational action (downward impulse), or to increase it if the action was below the ex-post rational action (upward impulse). The size of an impulse is equal to the value of lost profit, multiplied by parameter θ if the profit falls below the reference level. It is commonly assumed that $\theta = 2$, in accordance with the prospect theory (Kahneman and Tversky, 1979), but we will permit a wider range of parameter values, including $\theta = 1$ (no loss aversion). IBE is defined as a point at which expected downward and upward impulses are equalized, therefore it is the rest point of dynamics modeled by the learning direction theory (Selten and Stoecker, 1986). The reliance on ex-post rationality makes IBE especially relevant for the current study because the treatment manipulations should change the likelihood of the two types of impulses and the ease with which the ex-post rational action can be identified. IBE has also been found to explain deviations from theoretical predictions in other settings, e.g. overbidding in auctions (Selten et al., 2005; Ockenfels and Selten, 2005) and "pull to the center" bias in the newsvendor game (Ockenfels and Selten, 2014), therefore it could potentially explain over-investments in contests. Since IBE has never been applied to contests, we derive the predictions for all the treatment variations in Appendix C.

Quantal response equilibrium (McKelvey and Palfrey, 1995) assumes a probabilistic choice rule, but requires consistency between actions and beliefs, just as in the Nash equilibrium. In particular, if Δ is a player's belief about the distribution of opponent's action, the decision utility for action s^k is calculated as the sum of expected payoff and a noise term:

$$u_i(s^k, \Delta) = \pi(s^k, \Delta) + \varepsilon_k$$

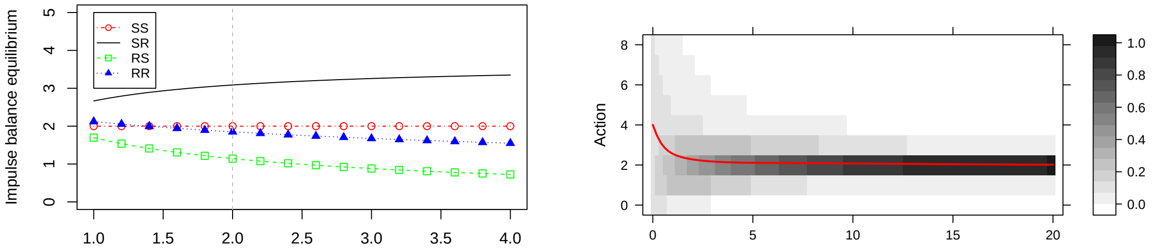
If the noise term is independently drawn from an extreme value distribution with a parameter $\lambda > 0$, the probability to choose action s^k is calculated as:

$$P(s^k) = \frac{e^{\lambda \pi(s^k, \Delta)}}{\sum_m e^{\lambda \pi(s^m, \Delta)}}$$

The logit QRE is a distribution Δ that satisfies $P = \Delta$ (McKelvey and Palfrey, 1995), because beliefs in equilibrium must be correct. QRE approaches Nash equilibrium as the precision parameter λ approaches infinity.

IBE and QRE predictions calculated for the parameters used in the experiment are displayed in figure 4. IBE predicts over-investments in SR and under-investments in RS because the intensity of regret is highest when players could have received the prize by a slight increase

between exploration and exploitation, and no reference point). We do not use the term "fictitious play" to avoid confusion with models based on explicit updating of beliefs about opponent's type.



(a) Impulse balance equilibrium, θ

(b) Quantal response equilibrium, λ

Figure 4: QRE and IBE predictions for each treatment. Dashed gray line in panel (a) marks the standard parameter value. Panel (b) shows the density of each action and the average action in the QRE, identical for all treatments.

in contest investment (in SR) or non-contest investment (in RS). This result holds for any value of θ, but higher loss aversion increases the intensity of impulses and the gap between SR and RS. QRE predicts uniform play when players are completely insensitive to payoff differences ($\lambda = 0$) and increased concentration around RNNE as sensitivity increases.

4.3 Nash Equilibrium with Expected Utility Maximization

RNNE is calculated assuming that expected utility is equal to the expected payoff and subjective probabilities are equal to objective probabilities. In this section, we relax these assumptions and study how the predictions change with the addition of social or risk preferences, non-monetary utility from winning and probability weighting.

Non-standard preferences are incorporated into the expected utility function by defining the lottery outcome $l \in \{1, \ldots, L\}$ for player i as $(\pi_i^l, \pi_j^l, W_i^l)$, where π_i^l is the payoff received by i, π_j^l is the payoff received by opponent j and W_i^l is the number of prizes that i has won in that round. Information about the opponent's payoff and the number of prizes won is required to calculate social preferences and non-monetary utility of winning.

The expected utility of player i is calculated as a weighted average of utilities at each outcome. Utilities are weighted using function $w(p_l)$, where p_l is the objective probability that outcome l will occur.

$$E[u_i] = \sum_{l=1}^{L} w(p_l) u_i(\pi_i^l, \pi_j^l, W_i^l)$$

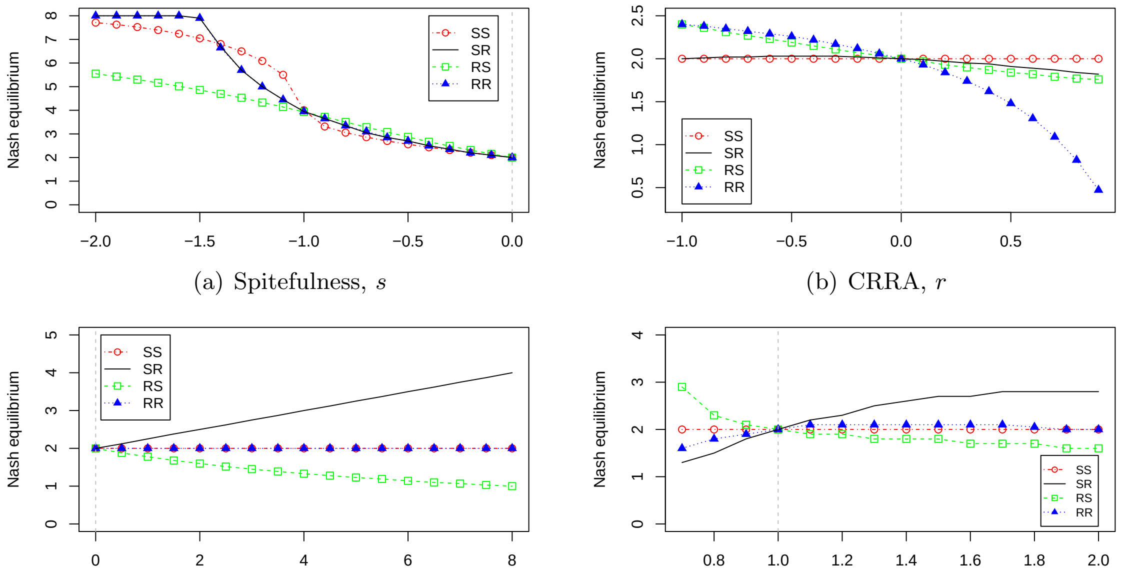
With standard preferences, $w(p_l) = p_l$ and $u_i(\pi_i^l, \pi_j^l, W_i^l) = \pi_i^l$, reducing expected utility to expected payoff.

Next, we specify the utility functions used to calculate utility at each outcome.

- **Social preferences.** We assume that in addition to own payoff, players care about opponent's payoff, weighted by s: preferences are altruistic if $s > 0$ and spiteful if $s < 0$.

$$u_i(\pi_i, \pi_j, W_i) = \pi_i + s\pi_j$$

- **Risk preferences.** We assume constant relative risk aversion (CRRA). Players are risk averse if $r > 0$ and risk seeking if $r < 0$.



(a) Spitefulness, s

(b) CRRA, r

(c) Non-monetary utility of winning, ω

(d) Probability weighting, β

Figure 5: Contest investments in Nash equilibrium for each treatment. Dashed gray line marks the standard parameter value.

$$u_i(\pi_i, \pi_j, W_i) = \frac{\pi_i^{1-r}}{1-r}$$

- **Non-monetary utility of winning.** We follow Sheremeta (2010) and assume that players who win receive additional utility ω, in addition to the monetary prize value.

$$u_i(\pi_i, \pi_j, W_i) = \pi_i + \omega W_i$$

- **Probability weighting.** We allow the weight placed on outcomes to differ from the objective probabilities. Specifically, we assume rank-dependent weighting as in the cumulative prospect theory (Tversky and Kahneman, 1992) with the following weighting function:

$$w(p) = \frac{p^\beta}{(p^\beta + (1-p)^\beta)^{1/\beta}}$$

The models discussed above have been developed for other games, and applying them to our treatments requires additional assumptions. Appendix A explains these assumptions and lists all possible lottery outcomes in each treatment, necessary for expected utility calculation. Appendix B discusses alternative specifications of social preferences (inequality aversion) and risk preferences (constant absolute risk aversion).

Figure 5 shows Nash equilibrium predictions for a range of parameter values.[18] Nash

[18] All calculations are done for the parameters used in the experiment but allowing for a continuous strategy space. For the range of parameters that we consider, a pure strategy Nash equilibrium always exists. When multiple equilibria exist, we plot the average action in all equilibria.

Table 2: Predictions about the treatments in which RNNE choices will be observed. Reinforcement learning and IBE predictions depend on the availability of FPI.

Theory	SS	SR	RS	RR
RNNE	+	+	+	+
Spite, QRE	−	−	−	−
Risk seeking	+	−	−	−
NMU of winning, prob. weighting	+	−	−	+
Reinforcement, IBE	− (if no FPI) + (if FPI)	−	−	−

equilibrium with spitefulness ($s < 0$ in panel a) predicts over-investments in all treatments, and there is no predicted treatment difference for moderate levels of spitefulness. Risk-seeking preference ($r < 0$ in panel b) predicts very moderate levels of over-investments in SR, higher over-investments in RS and RR and RNNE play in SS, where the risk is not present. Non-monetary utility from winning ($\omega > 0$ in panel c) predicts above-RNNE contest investments in SR, below-RNNE contest investments in RS and RNNE play in SS and RR. Inverse S-shaped probability weighting ($\beta < 1$ in panel d), commonly observed in previous studies, predicts below-RNNE contest investments. Over-investments in SR could be justified by S-shaped weighting function ($\beta > 1$ in panel d). This type of weighting predicts small over-investments in RR and under-investments in RS.

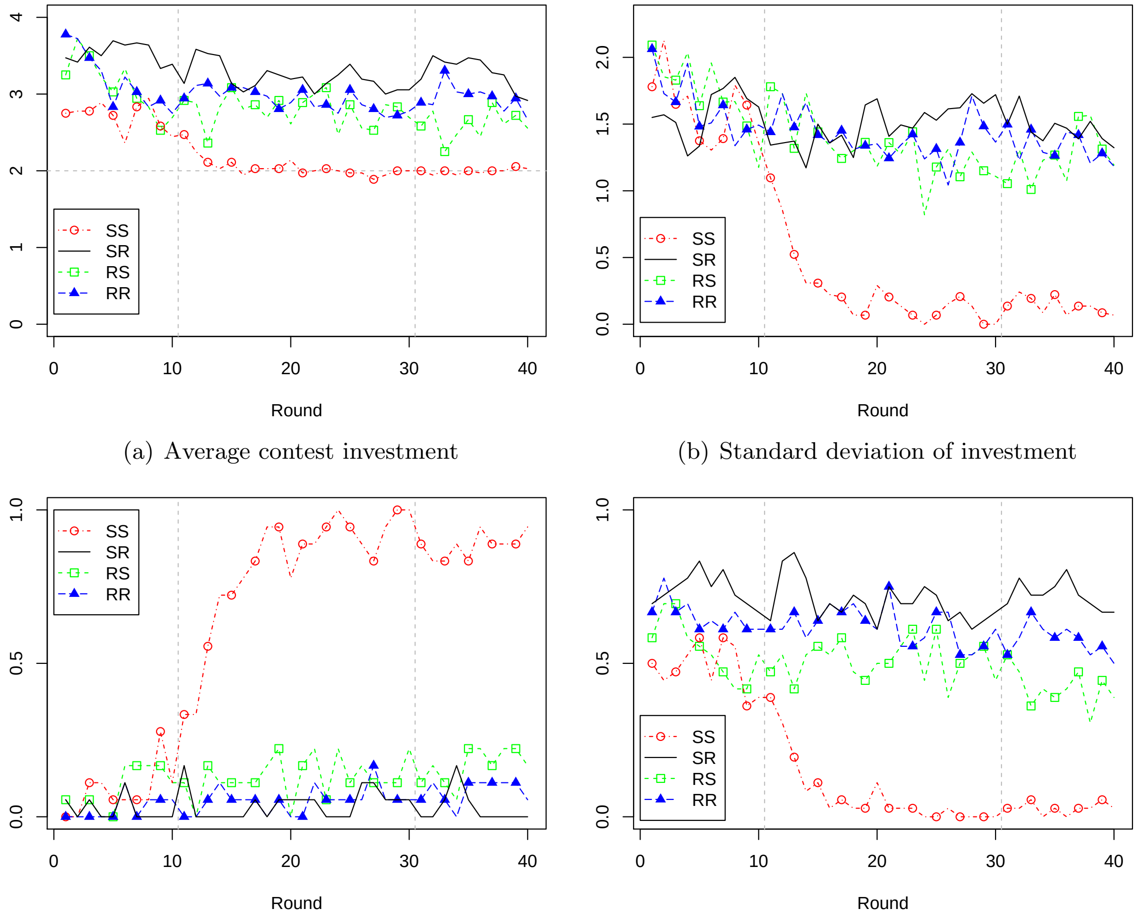
4.4 Summary of predictions

Specific predictions about behavior in each treatment depend on the parameter values, but broad predictions about RNNE play across treatments are robust to the model specification (see table 2 for a summary). Theories based on social preferences predict deviations from RNNE in all treatments because contest investments always reduce the expected earnings of the opponent. Theories based on risk preferences predict RNNE play in SS, which has no risk, and deviations from RNNE in all other treatments. RNNE maximizes ex-post payoffs only in SS, therefore IBE predicts RNNE in SS but not in the other three treatments. Theories based on the non-monetary utility of winning predict higher investments into the option with a probabilistic outcome. QRE depends only on the expected payoffs and therefore predicts identical deviations from RNNE in all treatments.

A prediction that FPI increases RNNE rates is made only by reinforcement learning (because FPI increases the quantity and quality of feedback) and by IBE (because FPI makes it easier to find the ex-post rational action and increases the intensity of impulses). Other theories implicitly assume that common knowledge of the payoff function is sufficient to calculate payoffs, therefore the provision of FPI has no additional effect.

5 Results

We find that over-investments and over-spreading disappear when FPI is introduced, but only in the SS treatment. The average contest investment and the standard deviation are shown



(a) Average contest investment

(b) Standard deviation of investment

(c) Fraction of pairs in RNNE

(d) Fraction of dominated strategies

Figure 6: Dynamics in each treatment. FPI was available from round 11 to round 30.

in panels (a) and (b) of figure 6.[19] Dispersion and average investments are high in the first 10 rounds, but sharply decrease in SS following the introduction of FPI in round 10. In the other three treatments, high dispersion and investments persist throughout the game. We test if these differences are significant by calculating the average and standard deviation of contest investments in the first 10 rounds and last 10 rounds for each matching group. Tables 10 and 11 in Appendix D show the two-sided p-values of Mann-Whitney U tests for all possible treatment comparisons. In the first ten rounds, the average investment is lower in SS than in SR ($p = 0.01$), and there are no significant differences between all other treatments. In the last ten rounds, both average investment and the standard deviation are significantly lower in SS than in each of the other three treatments (p-values at most 0.016).

A treatment difference in terms of average investments and dispersion does not indicate whether choices converge to the RNNE prediction. We therefore use two measures of convergence and compare them across treatments. First, we test whether the average investment is significantly different from the RNNE prediction. In the first 10 rounds, contest invest-

[19] Standard deviation is calculated separately for each matching group (i.e. using the decisions of 6 people in each group) and then averaged. See figure 14 in Appendix G for the evolution of the entire choice distribution.

ments are significantly above RNNE in all treatments (one sample two-tailed t-test p-values are at most 0.0055). In the last 10 rounds, investments are not different from RNNE in SS (two-sided p-value = 0.8125) but they are above RNNE in the other three treatments (all p-values below 0.01). Second, we compare the treatments in terms of the fraction of pairs in each matching group who choose exactly the RNNE action profile. Panel (c) in figure 6 indicates that convergence occurs only in SS and statistical tests confirm this observation (see table 12 in Appendix D for all pairwise treatment comparisons). RNNE rates are low in all treatments at the start of the game (the only significant difference is found between RS and SR, $p = 0.03$). In rounds 10-30, RNNE rates reach 100% in SS but remain at most 20% in SR, RS and RR. In the last block, the difference between SS and the other three treatments is highly significant (p-values at most 0.004).

Deviations from RNNE could be rationalized by non-equilibrium beliefs, but dominated strategies should not be chosen even under much weaker assumptions on rationality. It is known that dominated strategies are commonly chosen in contests (Masiliūnas et al., 2014), and we test whether their frequency decreases when players receive good feedback on mistakes. Panel (d) in figure 6 shows that there is no significant difference in terms of dominated strategies in the first ten rounds (the only significant difference is between SS and SR, $p = 0.01$). The gap between treatments appears after round 10, when the fraction of dominated strategies shrinks to 0% in SS but remains above 50% in the other treatments. Table 13 shows that the difference between SS and the other three treatments in the last ten rounds is highly significant ($p = 0.004$), although the difference between RS and SR becomes significant as well ($p = 0.02$).

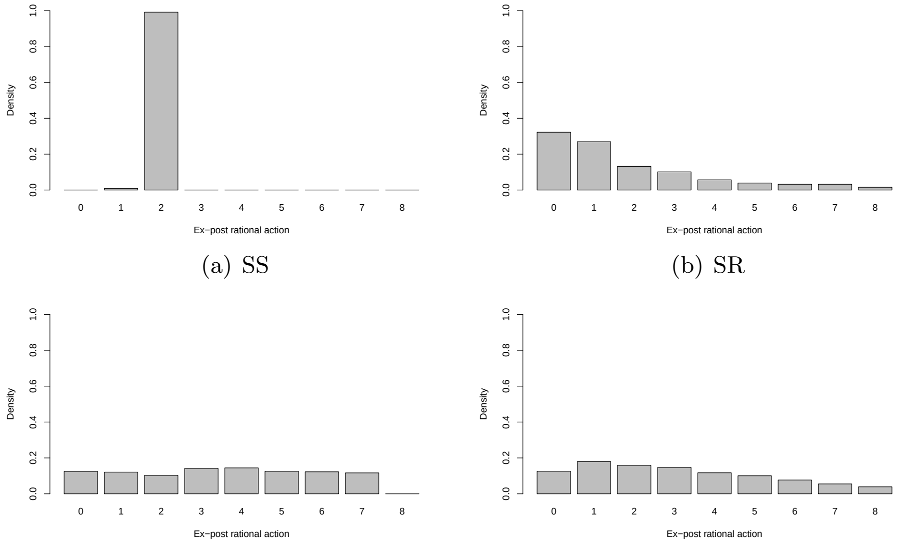
Within each treatment, we can identify what has been learnt by comparing choices in the first 10 rounds to the last 10 rounds. Both blocks are comparable because they have the same number of rounds and are played without FPI. Contest investments significantly decrease in SS (paired t-test p = 0.0034), and, to a lesser extent, in RS (p = 0.0375). Figure 16 in Appendix G compares investment distributions in the first ten and the last ten rounds. The fraction of RNNE choices increases from 18% to 32% in RR (two-tailed t-test p-value = 0.0072), from 25% to 42% in RS (p = 0.0171), from 29% to 94% in SS (p = 0.0002), but remains at 17% in SR (p = 0.5375). An increase in the explanatory power of RNNE is therefore highest in SS, lower in RR and RS, and not observed in SR. Further evidence about what participants learn in the experiment comes from answers about the action that would be recommended to a friend in a hypothetical future experiment. RNNE action is recommended by 81% of participants in SS, 44% in RS, 36% in RR and only 8% in SR (see figure 17 in Appendix G for more details). The difference between SS and the other three treatments is highly significant (test of proportion p-value is at least 0.0008). Both measures indicate that learning is strongest in SS and weakest in SR.

Convergence in SS occurs once FPI is introduced, but it is possible that it would have occurred even in the absence of FPI. We did not run a treatment in which FPI was withheld in all rounds, because our main interest lies in the treatment effect in the first 10 rounds (without FPI) and the last 10 rounds (after experience with FPI). But we can address this question by comparing our data to previous studies that ran treatments equivalent to SS without FPI (Masiliūnas et al., 2014, Chowdhury et al., 2014, Fallucchi et al., 2013). Appendix E shows that the rates of deviations from RNNE in the first 10 rounds of SS are similar to the rates found in previous studies, but a subsequent convergence occurs only in SS.

We find that the average contest investment in SS is significantly below the other three treatments, and not different from RNNE prediction. It seems that no theory besides reinforcement learning can explain this result. Social preferences never predict investments to be lower in SS than in the other three treatments, risk-seeking preferences correctly predict the difference between SS and the other three treatments, but the effect size predicted by CRRA utility function is much smaller than the difference observed in experiments. The effect of risk preferences should also appear already in the first block. Non-monetary utility of winning, probability weighting and IBE predict that if investments are lower in SS than in SR, then in RS they should be even lower. Experiments fail to find this effect, and instead show that the important factor is the presence rather than the nature of the payoff risk. QRE does not predict any treatment difference. However, the comparison of theories based solely on the predicted rank of treatments is very crude and excludes the possibility that multiple factors interact. Section 5.2 will test whether the results can be explained by a richer model that combines learning and a utility function with multiple elements.

5.1 Feedback and adaptation

We have shown that choices converge to RNNE in SS, but not in the other treatments. To understand the mechanism behind the treatment effect, we compare the obtained feedback and its effect on subsequent choices. This subsection evaluates only the response to feedback from previous period; the next subsection will estimate models that take into account the entire path of play.



(a) SS

(b) SR

(c) RS

(d) RR

Figure 7: Distributions of ex-post payoff-maximizing actions in rounds with FPI.

To measure what players could learn from the game, for each round we calculate the action that would have provided the highest payoff.[20] We use data from rounds with FPI, which allows the ex-post payoff-maximizing action to be known. Figure 7 shows that the

[20] We calculate the payoffs associated with each action by holding constant opponent's action and the draw of a random number, used to determine the lottery outcome. These payoffs are identical to those observed by participants. When the payoff-maximizing action is unique, it receives a weight of 1. When multiple actions generate the same maximal payoff, the weight is divided equally among those actions.

RNNE action ($c_i^* = 2$) almost always generates the highest payoff in SS, but not in the other treatments. In SR, payoffs are typically maximized by investing 0 or 1, which are optimal when winning is either impossible or guaranteed by choosing the smallest investment level (for more details, see Appendix C). In RS and RR, all actions maximize payoffs with similar frequency, thus the observed payoff information does not allow to identify the expected payoff-maximizing action. We conclude that the feedback from a single round in SR, RS and RR treatments reveals little information that could improve the quality of subsequent decisions.

Next, we study how the observed feedback affects the direction in which participants subsequently adjust their choices. We compare the predictions the following theories:

- **Cournot best-response** predicts that players choose the action that maximized the expected payoffs in the previous round. Cournot best-response would converge to RNNE in all treatments.

- **Imitate-the-best** predicts that players adapt towards the action of the player who received highest earnings. In a two player game, players either change actions in the direction of opponent's action (if opponent received a higher payoff), or make no change. In SS, imitate-the-best would converge to the relative payoff maximization point (Fallucchi et al., 2013), equal to 4 with the parameters of this experiment.

- **"Chasing" hypothesis** predicts that players adapt towards the action that provided the highest foregone payoff (Ert and Erev, 2007). This adaptation rule ignores the magnitude of payoffs and therefore may converge to the action that outperforms other actions most of the time, instead of the one that maximizes expected payoffs.[21] In contests, "chasing" converges to RNNE only in SS, while in other treatments it oscillates depending on the lottery outcomes. We expect that "chasing" will be used only in rounds with FPI. Since we have information about which FPI was actually observed, we will also test if players choose actions with the highest payoff from the set of actions with observed FPI.

Theories are compared by estimating whether the change in investment can be explained by observed feedback. We use a model adapted from Huck et al. (1999):

$$c_i^t - c_i^{t-1} = \beta_0 + \beta_{RP}(c_{RP}^{t-1} - c_i^{t-1}) + \beta_{OP}(c_{OP}^{t-1} - c_i^{t-1}) + \beta_{EP}(c_{EP}^{t-1} - c_i^{t-1}) + \beta_I(c_I^{t-1} - c_i^{t-1}) + \varepsilon_i^t$$

where c_{RP}^{t-1} is the action with the highest *realized payoff* in round $t-1$; c_{OP}^{t-1} is the action with the highest *observed realized payoff* in round $t-1$; c_{EP}^{t-1} is the action that would have maximized the *expected payoff*; c_I^{t-1} is the action chosen by the player with the highest payoff. The "chasing" hypothesis predicts adaptation towards c_{OP}^{t-1}, Cournot best-response predicts adaptation towards c_{EP}^{t-1} and imitate-the-best predicts adaption towards c_I^{t-1}.

We start by estimating the parameter values from rounds in which FPI was available. First, we calculate the action that maximized foregone payoffs, regardless of whether they were observed or not.[22] Table 3 displays the estimated parameter values from all treatments

[21] It has been shown that "chasing" can reduce efficiency if there is another option that has higher long-run benefits (Otto and Love, 2010) or if the option that is usually better has a lower expected value (Yechiam and Busemeyer, 2006).

[22] If several actions provide the same maximal payoff, we calculate their average. We will use the average value in all specifications so as not to discard any observations, but the results are very similar if observations with multiple payoff-maximizing actions are discarded.

Table 3: Random-effects GLS regression with a matching-group-specific random component. Independent variable is the change in contest investment. Standard errors are clustered on a matching group level.

	(1) Rounds with FPI	(2) Rounds with FPI	(3) Rounds without FPI
β_{RP}	0.0586***	0.00699	0.0187
	(3.94)	(0.41)	(1.02)
β_{OP}		0.163***	
		(6.12)	
β_{EP}	0.223***	0.225***	0.327***
	(5.17)	(5.30)	(9.61)
β_I	0.166***	0.160***	0.206***
	(5.99)	(5.85)	(5.59)
Constant	0.161***	0.146***	0.307***
	(3.35)	(3.25)	(4.61)
Overall R^2	0.205	0.228	0.244
N. of observations	2736	2736	2592

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

pooled together. In the first model, all coefficients are significantly higher than zero, but the largest coefficient belongs to the expected payoff-maximizing action. If parameters are estimated separately for each treatment, adjustment towards expected payoff-maximizing action is significant in all treatments, imitation is not significant in RR and SS, and adjustment towards foregone payoff-maximizing action is not significant in RR.

Since we have information about which foregone payoffs are observed, we test the hypothesis that players adapt in the direction of the action with the highest observed payoff. Model (2) shows that this variable (β_{OP}) is highly significant, while β_{RP} is no longer significant. Results are very similar if we set $\beta_{RP} = 0$ or if we remove rounds in which FPI was not observed (although the p-value of β_{OP} increases to 0.063), or if we additionally remove rounds in which multiple actions provide identical maximal payoffs.

Next, we look at rounds in which FPI was not available (1-10 and 31-40). The software calculated foregone payoffs in all rounds, but since this information was withheld, we expect it to play no role in the adjustment process. If the parameter is still significant, adjustment towards the foregone payoff-maximizing action would be caused by some other reason. Column 3 in table 3 shows that it is not: the estimated parameter value is small and insignificant, while the parameters for imitation and Cournot-best-response are highly significant. When we estimate the model separately for each game and either the first or the last block of 10 rounds, we find that foregone payoffs do no play any role in all games except for SS, where its predictions coincide with those of expected payoff maximization. Adaptation towards expected payoff-maximizing action is highly significant in all treatments and blocks, while imitation is significant in the first block for SS, SR and RS treatments (in other three treatments p-value is at most 0.006), but not significant in the last block (p-value at least 0.152). This decrease over time suggests that imitation is replaced by more rational adaptation rules once players gain more experience.

Overall, we find that the primary difference between SS and the other three treatments lies in the observed feedback. We further study the joint effect of observed feedback and the subsequent choice in Appendix F, finding evidence that participants choose the action that was seen providing the highest payoff.

5.2 Learning and non-standard preferences

We find that RNNE is played only in the treatment with no payoff risk and with FPI, just as predicted by reinforcement learning simulations. It remains to be tested whether reinforcement learning can explain the entire adjustment process in each treatment. We are also interested in comparing reinforcement learning to a model with non-standard preferences and probability weighting. The comparison would not be fair if we modeled preferences using a static solution concept, as in section 4.3, therefore we use a belief learning model extended with non-standard preferences and probability weighting. We then combine the two models using an extension of EWA. This estimation strategy identifies how much the addition of non-standard preferences improves the fit of a belief learning model, whether it fits better than reinforcement learning, and how much each of these models contributes to the fit of the most general EWA model.

5.2.1 Adaptation rules

Reinforcement learning

Denote the action space by $S \equiv \{s^1, s^2, \ldots, s^K\}$, and the action of player i in round $t \in [1, \ldots, 40]$ by $s_i(t) \in S$. The reinforcement learning model follows Roth and Erev (1995) and Erev and Roth (1998) in assuming that the attraction of action s^k, for $k \in \{1, 2, \ldots, K\}$, is a weighted sum of the payoff flow $\pi_k(t)$ that was generated or would have been generated by playing s^k:

$$A_k(t) = \phi A_k(t-1) + \delta_k(t)\pi_k(t) \tag{1}$$

Parameter $\phi \in [0, 1]$ determines the rate at which old payoff information is discounted. If $\phi = 1$, all past payoffs receive the same weight. If $\phi = 0$, only the most recent payoff is taken into account. Variable $\delta_k(t)$ does not appear in reinforcement learning models (e.g. Erev and Roth, 1998) and is added to allow learning from FPI. It attains one of the following values:

$$\delta_k(t) = \begin{cases} 1 & \text{if } s_k \text{ was chosen in round } t, \\ \delta_o & \text{if } s_k \text{ was not chosen, but its FPI was observed,} \\ \delta_a & \text{if } s_k \text{ was not chosen, its FPI was available but not observed,} \\ \delta_u & \text{if } s_k \text{ was not chosen and FPI was unavailable.} \end{cases} \tag{2}$$

Chosen actions are reinforced using realized payoffs and a weight of $\delta_k(t) = 1$. Unchosen actions are reinforced using foregone payoffs and a weight $\delta_k(t) \in [0, 1]$.[23] We allow the weight to depend on the type of FPI (observed, available but not observed, unavailable). If δ_o, δ_a or $\delta_u = 0$, that class of FPI is ignored. If δ_o, δ_a or $\delta_u = 1$, that class of FPI receives the same weight as the realized payoff.

[23] A similar approach was used to model reinforcement learning with FPI by Yechiam and Rakow (2012) and Yechiam and Busemeyer (2006).

Belief learning with non-standard preferences

The second learning model assumes learning from observed opponent's actions. Denote the strategy chosen by i's opponent by $s_{-i}(t)$. Using an indicator function $I(s^k, s_{-i}(t))$, which equals 1 if the opponent's chosen action in round t is s^k, and 0 otherwise, define the weighted frequency of strategy k played by i's opponent up to time t by $N^k(t) = \phi N^k(t-1) + I(s^k, s_{-i}(t))$. If $\phi = 1$, $N^k(t)$ reduces to the total number of times that action s^k was played by the opponent. Beliefs are normalized to attain values between 0 and 1 using an experience weight $N(t) = \phi N(t-1) + 1$. The belief about the probability that i's opponent will play s^k in round t is then calculated by:

$$b_k(t) = \frac{N^k(t)}{N(t)} = \frac{\sum_{u=1}^{t} \phi^{u-1} I(s^k, s_{-i}(t+1-u))}{\sum_{u=1}^{t} \phi^{u-1}} + \phi^t N^k(0) \tag{3}$$

Initial belief $N^k(0)$ represents prior experience, and $N(0)$ is its weight.

Attractions used to calculate choice probabilities are calculated as a belief-weighted average of expected utility:[24]

$$A_k(t) = \sum_{j=1}^{K} E[u_i(s^k, s^j)] b_j(t) \tag{4}$$

If expected utility equals expected payoff, the belief learning model reduces to weighted fictitious play by Cheung and Friedman (1997). The extension to expected utility permits non-standard preferences and learning to be modeled in one unified framework.

Generalized Experience-Weighted Attraction Learning

Next, we introduce a generalized experience-weighted attraction (EWA) learning model (Camerer and Ho, 1999), which combines the two learning models described above. In EWA, belief learning is implemented through sensitivity to FPI. Instead of forming explicit beliefs from the weighted path of observed choices and calculating expected payoffs conditional on these beliefs, players are directly calculating the weighted average of foregone payoffs in all previous rounds, which depend on the path of opponent's choices. The equivalence between belief and reinforcement learning exists only if the foregone payoffs of all actions receive the same weight, and if there is no payoff risk. If payoffs are risky, expected foregone payoffs must be used for EWA to be equivalent to weighted fictitious play. Generalized EWA (adapted from Shafran, 2012) uses both the expected payoffs, needed for belief learning, and the realized and foregone payoffs, needed for reinforcement learning.

The updating rules are governed by the experience weight $N(t)$ and attractions $A_k(t)$:

$$N(t) = \phi(1-\kappa)N(t-1) + 1 \tag{5}$$

$$A_k(t) = \frac{\phi N(t-1) A_k(t-1) + \gamma E[u(s^k, s_{-i}(t))] + (1-\gamma)\delta_k(t)\pi_k(t)}{\phi N(t-1)(1-\kappa) + 1} \tag{6}$$

This version of generalized EWA extends Shafran (2012) by allowing the weight of foregone payoffs to depend on the observability ($\delta_k(t)$ is calculated using equation 2) and by permitting

[24] Expected utility of player i from playing strategy s^k and opponent playing s^j is denoted by $E[u_i(s^k, s^j)]$.

expected utility. Generalized EWA reduces to standard EWA if there is no payoff risk (as in SS), $\gamma = 0$ and $\delta_o = \delta_a = \delta_u$. It reduces to cumulative reinforcement learning (equation 1) if $N(0) = 1$, $\gamma = 0$ and $\kappa = 1$. It reduces to averaged reinforcement learning if $N(0) = \frac{1}{1-\phi}$, $\gamma = 0$ and $\kappa = 0$. It reduces to weighted fictitious play (equation 3) if $\gamma = 1$, $\kappa = 0$ (see Camerer and Ho, 1999, and Shafran, 2012, for a proof).

We set $N(0) = 1$, as is common in the literature (see Camerer, 2003). Initial attractions $A_k(0)$ were chosen to maximize the likelihood of first round observations (Ho et al., 2008).[25]

The separation between expected and realized payoffs permits the restriction to either reinforcement or belief learning using only the γ parameter, while δ measures purely the difference between realized and foregone payoffs, manipulated in the experiment. In the standard EWA, δ captures both the difference between belief and reinforcement, and the sensitivity to FPI.

In all models, the probability to choose action k is calculated using a logistic choice rule:

$$P_k(t) = \frac{e^{\lambda A_k(t-1)}}{\sum_{j=1}^{K} e^{\lambda A_j(t-1)}}$$

where parameter $\lambda \in [0, \infty)$ measures sensitivity to differences between attractions.

We estimate the parameters that maximize the following log-likelihood function:

$$LL = \sum_{i=1}^{96} \sum_{t=1}^{40} \log(P_{s_i(t)}(t))$$

A potential problem for generalized EWA is overfitting, as models with many parameters may fit well but make inaccurate predictions. For this reason, we estimate parameters using 2/3 of the observations (96 participants), and use the remaining 1/3 to evaluate the out-of-sample goodness of fit.[26] Since a higher number of parameters increases the log-likelihood, we measure the goodness of fit using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which penalize models for the number of parameters.[27]

Data for all estimations is pooled from all four treatments and all rounds. We do so to test if one set of parameter values can explain the differences between all treatments and information conditions. If parameters were allowed to vary by treatment, we would likely observe overfitting, as models would predict less noise and lower strength of non-standard preferences in SS compared to the other tree treatments.

Estimations were performed using quasi-Newton and derivative-free optimization routines with various starting values and the following constraints: $\lambda > 0$, $\phi, \kappa, \delta, \gamma \in [0, 1]$, $s \in [-5, 5], r \in [-2, 1), \omega \in [0, 16], \beta \in [0, 3]$. Standard errors were calculated from the variance-covariance matrix, estimated from the numerical approximation of the Hesssian matrix.

[25] If f^k is the frequency of strategy s_i^k in round 1, $A_k(0) = \log(f^k)/\lambda$ if $f^k > 0$ and $A_k(0) = 0$ otherwise. It is straightforward to verify that these initial attractions generate the appropriate initial choice frequencies, but they are not unique, as the exponential choice rule is invariant to adding a constant to all attractions. Alternatively, we could have estimated the initial attractions as separate parameters, but additional 8 parameters would make model fitting very challenging.

[26] We have 6 independent matching groups in each treatment, so we use the first 4 groups for estimation (i.e. the four groups that were run in earlier sessions) and the last 2 for control.

[27] $\text{AIC} = -2\log(\mathcal{L}) + 2k$, $\text{BIC} = -2\log(\mathcal{L}) + k\log(N)$, where k is the model degrees of freedom and N is the number of observations.

5.2.2 Estimation results

Since generalized EWA nests belief and reinforcement learning, we estimate all models using equation (6) and obtain belief or reinforcement learning by appropriately constraining the parameter values.

Reinforcement learning

Reinforcement learning is obtained by setting $\gamma = 0$. We allow κ to vary between 0 and 1, to allow for both averaging and cumulation of attractions, although in all models the estimated value of κ is equal to 1, reducing the updating rule to equation (1).

We estimate and compare the fit of five reinforcement learning models that differ in assumptions about how FPI is used to update attractions:

1. FPI is ignored, so only realized payoffs receive a positive weight. This assumption is made in classical reinforcement learning models, e.g. Erev and Roth (1998).

$$\delta_a = \delta_o = \delta_u = 0$$

2. All FPI receives the same weight. This assumption is made in studies that do not provide explicit FPI. In the treatment with no payoff risk, this model is equivalent to the standard EWA model, and δ measures sensitivity to foregone expected payoffs.

$$\delta_a = \delta_o = \delta_u \in [0, 1]$$

3. Available FPI receives a different weight than unavailable FPI. This is the approach taken by studies that explicitly provide FPI (Yechiam and Rakow, 2012, Yechiam and Busemeyer, 2006).

$$\delta_o = \delta_a \in [0, 1], \delta_u \in [0, 1]$$

4. Only observed FPI receives a positive weight. This specification is unique to our study because we have information about the FPI that was observed.

$$\delta_o \in [0, 1], \delta_a = \delta_u = 0$$

5. Each type of FPI receives a different weight. This is the most general model, allowing attractions to be affected differently by each type of foregone payoffs.

$$\delta_o \in [0, 1], \delta_a \in [0, 1], \delta_u \in [0, 1]$$

Table 4 lists the estimated parameter values and the goodness of fit for all five reinforcement learning models. The standard model (1) that ignores FPI has a much worse fit than all other models. If all three types of FPI are treated the same way (model 2), the estimated weight placed on foregone payoffs is 0.14, much smaller than the weight of 1 that realized payoffs receive. Allowing a different weight for available but unseen FPI (model 3) does not improve the fit, as the estimated value of δ_a remains unchanged. If it is assumed that players

Table 4: Estimated parameter values and goodness of fit for reinforcement learning models in which (1) FPI is ignored, (2) all FPI is perceived the same, (3) available FPI receives a different weight than unavailable FPI, (4) only observed FPI receives a positive weight, (5) each type of FPI receives a different weight. Standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)
λ	0.15	0.19	0.19	0.16	0.19
	(0.015)	(0.023)	(0.023)	(0.019)	(0.024)
ϕ	0.96	0.94	0.94	0.95	0.94
	(0.0038)	(0.0042)	(0.0042)	(0.0041)	(0.0042)
κ	1	1	1	1	1
	(0.078)	(0.11)	(0.11)	(0.09)	(0.11)
δ_u	[0]	0.14	0.14	[0]	0.14
		(0.011)	(0.020)		(0.020)
δ_a	[0]	[δ_u]	0.14	[0]	0.11
			(0.016)		(0.019)
δ_o	[0]	[δ_u]	[δ_a]	0.30	0.24
				(0.037)	(0.036)
LL (in)	-5568.37	-5481.71	-5481.71	-5534.49	-5476.81
AIC	11142.73	10971.41	10973.41	11076.97	10965.61
BIC	11161.49	10996.43	11004.68	11101.99	11003.13
LL (out)	-2623.07	-2559.84	-2559.81	-2581.60	-2552.30

react only to observed FPI, while unseen FPI is ignored (model 4), the fit is worse compared to the two previous models. Allowing different weights for each type of FPI (model 5) has the best fit, both in-sample and out-of-sample, even when accounting for the additional number of parameters (although model (2) does better in terms of BIC, which penalizes additional parameters more than AIC). The estimation shows that the weight placed on observed FPI is twice higher than the weight placed on unobserved FPI, but still much lower than the weight placed on realized payoffs. The estimated value of δ_o is similar to that in model (4). We conclude that information about foregone payoffs affects the adaptation process and including it improves the fit.

Belief learning with non-standard preferences

We estimate the belief learning models with the utility functions introduced in section 4.3. Additionally, non-monetary utility from winning is allowed to depend on the type of contest. Lottery outcome $l \in \{1, 2, \ldots, L\}$ is therefore defined as $(\pi_i^l, \pi_j^l, C_i^l, NC_i^l)$, where $C_i^l = 1$ if i received the contest prize and $NC_i = 1$ if i received the non-contest prize (and 0 otherwise). The payoffs of i and j are denoted by π_i and π_j. Utility at each outcome is calculated as:

$$u_i(\pi_i, \pi_j, C_i, NC_i) = \frac{\pi_i^{1-r}}{1-r} + \omega_c C_i + \omega_{nc} NC_i + s\pi_j \tag{7}$$

where s is the social preference parameter, r is the risk aversion coefficient assuming CRRA, ω_c is non-monetary utility from winning the contest prize and ω_{nc} is non-monetary utility from winning the non-contest prize.

Objective probabilities of lottery outcomes are transformed to subjective weights using the cumulative prospect theory weighting function (Tversky and Kahneman, 1992):

$$w(p) = \frac{p^\beta}{(p^\beta + (1-p)^\beta)^{1/\beta}} \tag{8}$$

If the probability of lottery outcome l is p_l, the expected utility is calculated as:

$$E[u_i] = \sum_{l=1}^{L} w(p_l) u_i(\pi_i^l, \pi_j^l, C_i^l, NC_i^l) \tag{9}$$

Equation (9) is used to calculate expected utilities in equation (6). Parameters of the utility function are estimated jointly with the parameters of the learning model. The drawback of using expected utility instead of expected payoffs is that the values of parameters such as λ are no longer comparable across specifications. To preserve comparability, we multiply the utilities by a scalar, keeping the average utility from uniform choices identical to the average payoff from uniform choices. Since we use EWA formulation, the belief learning model is comparable to reinforcement learning, having identical initial attractions and the same flexibility in aggregating past payoff information. The sole difference is the use of expected utility instead of realized payoffs.

Table 5: Estimated parameter values and goodness of fit for belief learning models with (6) standard preferences, (7) social preferences, (8) risk preferences, (9) non-monetary utility of winning, (10) probability weighting and (11) all preferences combined. Standard errors are in parentheses.

	(6)	(7)	(8)	(9)	(10)	(11)
	NO	SOC	RISK	NMU	PW	ALL
λ	0.06	0.78	0.064	0.22	0.064	0.45
	(0.01)	(0.011)	(0.012)	(0.051)	(0.011)	(0.028)
ϕ	0.97	1	0.93	0.91	0.97	1
	(0.0076)	(0.034)	(0.011)	(0.011)	(0.008)	(0.049)
κ	1	0.03	1	0.82	1	0.043
	(0.086)	(0.01)	(0.12)	(0.16)	(0.089)	(0.025)
s	[0]	-0.27	[0]	[0]	[0]	-0.64
		(0.01)				(0.041)
r	[0]	[0]	-0.80	[0]	[0]	-0.80
			(0.061)			(0.011)
ω_c	[0]	[0]	[0]	3.78	[0]	4.82
				(0.18)		(0.63)
ω_{nc}	[0]	[0]	[0]	0	[0]	0
				(0.13)		(0.18)
β	[1]	[1]	[1]	[1]	1.29	1.53
					(0.03)	(0.052)
LL(in)	-6889.65	-6731.22	-6776.60	-6605.92	-6826.78	-6552.2
AIC	13785.3	13470.45	13561.19	13221.85	13661.55	13120.41
BIC	13804.05	13495.46	13586.21	13253.11	13686.57	13170.43
LL (out)	-3306.82	-3321.89	-3302.38	-3357.66	-3312.05	-3357.03

Table 5 shows the estimated parameter values for belief learning. The baseline model (6) with standard preferences has a poor fit, much worse than the baseline reinforcement learning model. The next four models consider each of the four expected utility modifications. Each one fits better than the baseline model, even when accounting for the number of parameters. However, the out-of-sample fit is reduced in all but the risk preference model, a result driven by very small differences between SR, RS and RR in the out-of-sample groups. Of the four theories, non-monetary utility of winning fits the best on all measures, except for the out-of-sample fit, where it fits the worst. All parameter values are of the magnitude that could explain over-investments in the standard contest: players are anti-social, risk-seeking, receive non-monetary utility from winning the contest prize, but no non-monetary utility from winning the non-contest prize, and probability weighting is S-shaped. When all non-standard preferences are estimated jointly (model 11), the estimated values of preference parameters remain very similar. The model that combines all four components of the utility function fits better than other models, although still worse than the baseline reinforcement learning model. The out-of-sample fit of the combined model is among the worst and suggests overfitting.

Table 6: Estimated parameter values and goodness of fit for EWA models that combine belief and reinforcement learning. Standard errors in parentheses.

	(12)	(13)	(14)	(15)
λ	0.24	0.24	0.95	0.36
	(0.034)	(0.035)	(0.10)	(0.046)
ϕ	0.92	0.92	0.92	0.91
	(0.0051)	(0.0052)	(0.0059)	(0.0054)
κ	1	1	0.14	1
	(0.13)	(0.14)	(0.029)	(0.14)
δ_u	[0]	0	[0]	0
		(0.027)		(0.03)
δ_a	[0]	0	[0]	0
		(0.028)		(0.034)
δ_o	[0]	0.075	[0]	0.08
		(0.046)		(0.051)
s	[0]	[0]	-0.32	-0.069
			(0.48)	(0.0098)
r	[0]	[0]	-1.13	0.52
			(0.059)	(0.0037)
ω_c	[0]	[0]	16	1.25
			(1.34)	(0.13)
ω_{nc}	[0]	[0]	0.016	0
			(1.42)	(0.13)
β	[1]	[1]	1.53	1.35
			(0.072)	(0.059)
γ	0.25	0.25	0.21	0.51
	(0.014)	(0.020)	(0.017)	(0.026)
LL(in)	-5370.27	-5368.92	-5270.57	-5229.08
AIC	10748.55	10751.83	10559.14	10482.16
BIC	10773.56	10795.6	10615.42	10557.2
LL (out)	-2514.26	-2508.28	-2548.06	-2564.25

Generalized EWA

Next, we estimate the parameters for EWA model that combines reinforcement and belief learning. To maintain a meaningful interpretation of the γ parameter, we rescale the utilities from belief learning models, keeping average utilities equal to average payoffs. Table 6 shows the estimated parameter values. Model (12) is identical to generalized EWA (Shafran, 2012) and includes neither FPI nor the non-standard utility parameters. This model fits data better than any of the belief or reinforcement models, and the low estimated value of γ suggests that it is mainly driven by reinforcement rather than belief learning. The addition of foregone payoff parameters in model (13) improves the fit, but not by much, as the estimated δ parameters are all close to zero, except for the observed foregone payoffs. The additional utility function parameters in model (14) improve the in-sample fit further, although the out-of-sample fit is reduced. The estimated utility function parameters are of similar magnitude to the beliefs-only model parameters (table 5). Model (15) is the most general and has the best fit, both in terms of AIC and BIC. As in model (13), only the observed foregone payoff parameter δ_o receives a positive weight. The estimated parameters of the utility function parameters are similar to those in model (14), but their magnitude is reduced and risk seeking is replaced by risk aversion. Reduced strength of non-standard preferences is compensated by a higher weight assigned to belief learning (higher γ).

5.2.3 Goodness of fit

To understand why some models fit better than others, we compare the average contest investments and RNNE play rates observed in experiments (panels (a) and (c) in figure 6) to the predictions made by the best-fitting model in each class (the most interesting predictions are displayed in figure 8). The baseline reinforcement learning model (1) accurately predicts the difference of RNNE frequency between SS and the other three treatments, as well as the treatment order at the end of the experiment. Estimated parameter values of model (4) are close to model (1), therefore the two share a similar in-sample fit. Models (2), (3) and (5) make similar predictions because they add similar weights to all FPI, although learning in (5) is slightly faster due to the positive weight placed on unobserved FPI and a higher value of λ. All reinforcement models replicate the general features of the data, except for the under-predicted speed of convergence in SS.

The baseline belief learning model (6) fails to predict persistent over-investments and a difference between treatments, because expected payoffs are identical by design, and the distribution of beliefs differs only in SS. The addition of social preferences in model (7), risk preferences in model (8) or probability weighting in model (10) improves the fit by predicting slightly higher contest investments, but with no observable treatment difference. Model (9) assumes non-monetary utility of winning only for the contest prize, while the non-monetary utility from winning the non-contest prize is estimated to be zero. The model therefore predicts higher average contest investments and slightly lower RNNE frequency in SR and RR than in RS and SS. Model (11) combines all non-standard preferences and probability weighting: non-monetary utility increases the predicted contest investment in RR and SR, while probability weighting additionally increases the prediction in SR and decreases in RS. The full belief learning model correctly explains the treatment ranking in terms of average choices between SR, RS and RR, but fails to explain lower investments and higher RNNE rates in SS.

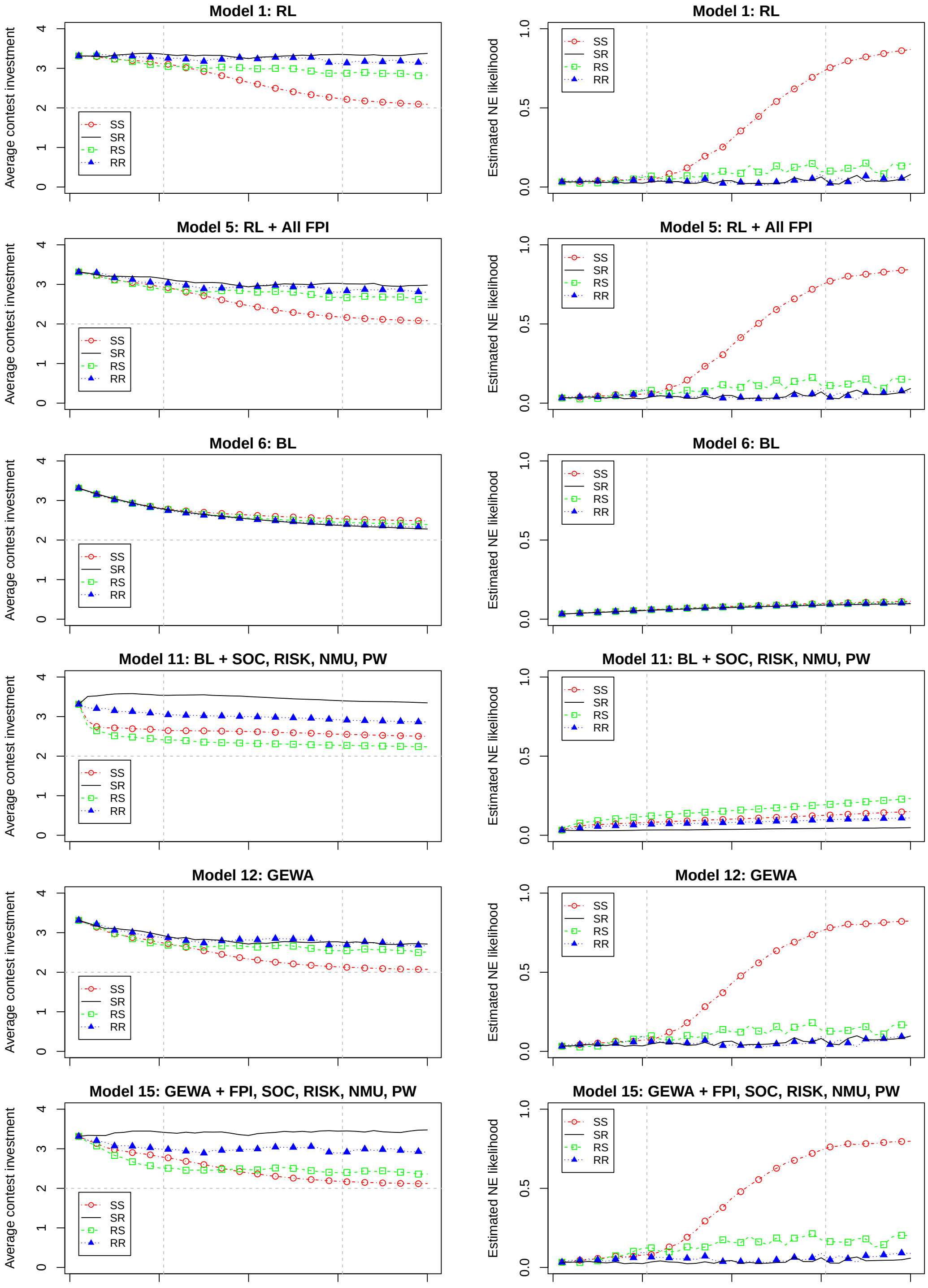


Figure 8: Average contest investments (left) and RNNE frequency (right), estimated using one-step-ahead predictions.

The generalized EWA model (12) that combines belief and reinforcement learning explains lower contest investments and higher RNNE rates in SS, compared to the other three treatments, but fails to predict a difference between SR, RS and RR. The difference between these three treatments is not predicted even with the added sensitivity to FPI (model 13), but it is predicted with the addition of non-standard preferences in models (14) and (15). Thus the main contribution of reinforcement learning lies in explaining lower contest investments and higher frequency of RNNE in SS (compare the belief model 11 to the full EWA model 15), while the main contribution of non-standard preferences and probability weighting is in explaining differences between SR, RR and RS (compare the reinforcement learning model 5 to the full EWA model 15). But while both models improve the fit, the contribution of reinforcement learning is much larger: it improves the total log-likelihood by 2116 points (difference between model 15 and 11), while the improvement from belief learning with non-standard preferences is only 236 points (difference between model 15 and 5).

Results from the estimations should be interpreted with caution. It is known that when learning models are misspecified, EWA fails to accurately identify populations from simulated data 50% of the time (Salmon, 2001). The problem is even more severe if the population is heterogeneous, as is often true in contest experiments; in that case, the estimated parameters could be very inaccurate (Wilcox, 2006). In particular, there tends to be a large downward bias in the estimation of the δ parameter, which in our model is interpreted as sensitivity to FPI. It is thus likely that participants are more sensitive to FPI than we found. The model that combines EWA with non-standard preferences is completely new, thus it is unknown to what extent and in which direction the estimated preference parameters could be biased. The benefit of our estimation strategy is that all models are estimated using data from all four treatments (in contrast to using one game, as is typically done in the literature), and models differ in how variation in risk is predicted to affect behavior. Thus even if the estimated parameter values were biased, the ranking of models in terms of the goodness of fit should remain similar. Ultimately, how accurate and useful are the results from the estimation exercise is an empirical question that should be addressed in future research, perhaps by testing how well the parameter values can be retrieved from simulated data.

6 Concluding remarks

We test whether deviations from Nash equilibrium occur because learning is subject to sampling error, caused by the low informational value of realized payoffs. We show that if participants were learning only from the information about realized payoffs, they would not find the payoff-maximizing action even if they were trying to do it. Experiments show that when the sampling error is reduced by removing payoff risk and providing foregone payoff information, the rates of risk-neutral Nash equilibrium play are dramatically increased. These results can be explained by payoff-based learning, but not by preference-based theories. The learning hypothesis is further supported by the findings that participants continue choosing RNNE actions even when foregone payoff information is removed and that most participants would recommend the RNNE action to their friends.

Our results complement rather than substitute the preference-based explanations proposed in the contest literature. The finding that participants respond to FPI as predicted by payoff-based learning does not mean that participants also use payoff-based learning in the absence of FPI. In fact, it is very likely that when learning is difficult, participants do not even attempt to learn and use other heuristics instead. This would explain why previous literature found that other-regarding preferences, non-monetary utility of winning and risk preferences play an important role in contest experiments. The question addressed here is why these factors play a large role, while expected payoff differences matter so little. Sheremeta (2016) shows that many measured personal characteristics affect behavior because they are correlated with impulsiveness. When viewed from the perspective of the dual processes model (Sloman, 1996), it is important to know why so many players use the impulsive type I rather than the deliberative type II system. We show that the answer may lie in the contest environment, which makes it difficult to utilize the type II system. We find that theories based on non-standard preferences and probability weighting can explain differences between treatments in which learning is difficult, but the effect of non-standard preferences disappears and players converge to RNNE when learning is made easier.

Variability in the quality of feedback can organize our data, but it can also organize data from other studies that manipulated payoff risk in contests. Chowdhury et al. (2014) find high Nash equilibrium rates only in the treatment with no payoff risk and quadratic investment costs. Quadratic costs increase the penalty of non-equilibrium choices; as a result, dominated actions would typically provide low payoffs and payoff-based learning would predict faster convergence. Masiliūnas et al. (2014) find that choices are close to theoretical predictions only when payoff risk is removed and opponents play a fixed action over time, manipulations that would dramatically increase the speed of payoff-based learning. Fallucchi et al. (2013) find increased Nash equilibrium rates only in the treatment with no payoff risk and no feedback on individual choices. Feedback about the choice and payoff of each opponent may nudge players to imitate-the-best, crowding out the forms of learning that would converge to the Nash equilibrium. We also find evidence for imitation at the start of the game, but its effect decreases in all treatments when the quality of feedback is increased.

Our main result is that nearly all participants behave as predicted by game theory when foregone payoffs are observed and there is no payoff risk. The sharp increase in RNNE play in this environment indicates that these manipulations made it very easy, perhaps even trivial, for participants to find the expected-payoff maximizing action. Identifying the conditions that facilitate rapid convergence is useful to understand the adaptation process in contests and games in general. However, in future research it would also be interesting to investigate weaker manipulations, which would make the ex-post rational action less obvious. Such research could more precisely measure the minimum amount of information needed for the convergence to Nash equilibrium to occur.

Important conclusions can also be drawn from the treatment differences that were not observed. We found no difference between the four treatments in the first ten rounds, when foregone payoff information was not available. This finding suggests that in the absence of foregone payoff information, over-investments persist and are robust to changes in payoff risk. This result is important because in practice contests differ in the degree and nature of payoff risk. Originally, the rent-seeking contest was used to study competition for monopoly rents (Tullock, 1980). In such winner-take-all markets the assumption of probabilistic prize allocation is reasonable because only one product can become the industry standard (e.g. either Blu-ray or HD DVD). But there are other situations in which the market is divided between competitors; for example, advertising and technical improvements may increase the market share, but not guarantee the entire market. Similarly, resources that are not spent on competition are rarely kept as cash but instead invested in other risky projects. Risk in these projects originates not from the uncertainty about the behavior of the competitors, but from the potential external technological or legal challenges. Our paper studies such variations in payoff risk and finds similar behavioral patterns in all treatments (as long as foregone payoffs are not displayed). This result gives confidence that the numerous results obtained using the standard Tullock contest success function can be extended to a more general framework. We also find that choices in the standard contest are robust to the provision of foregone payoff information. Throughout the experiment, participants have access to 200 payoff realizations, but the distribution of choices at the end of the game is almost identical to the initial distribution. This finding suggests that deviations from theoretical predictions in standard contests are unlikely to disappear when the number of repetitions is increased.

Standard solution concepts assume that players are fully informed about the incentive structure. However, in complex games boundedly rational participants may learn little from the game description, and instead rely on information acquired from experience. In such games, models that assume no prior information about the game structure may have higher explanatory power than those assuming full information. Experiments with individual choice tasks show that decisions made from experience are very different from decisions made from description, and a lot is known about how decisions are affected by realized payoff information (Wulff et al., 2018). We find that the quality of feedback is important in strategic situations and show how this information can be used to improve the accuracy of predictions. We hope that these results will make a step towards the development of a successful positive model of human behavior.

Our study contributes to the question of how the informational value of received payoff information influences the explanatory power of solution concepts in repeated games. Bereby-Meyer and Roth (2006) show that noisy payoffs slow down learning in prisoner's dilemma, and Shafran (2012) shows that stochastic payoffs slow down convergence in coordination games. We show that reduced noise dramatically improves Nash equilibrium play in Tullock contests. It remains to be tested if these findings can be extended to other games and if heterogeneity in the informational value of feedback can explain why convergence to equilibrium is observed in some games (e.g. competitive guessing game, Weber, 2003, race game, Gneezy et al., 2010) but not in others.

References

Baharad, E. and Nitzan, S. (2008). Contest efforts in light of behavioural considerations. *The Economic Journal*, 118(533):2047–2059.

Beggs, A. W. (2005). On the convergence of reinforcement learning. *Journal of Economic Theory*, 122(1):1–36.

Bereby-Meyer, Y. and Roth, A. E. (2006). The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation. *American Economic Review*, 96(4):1029–1042.

Bosch-Domènech, A. and Vriend, N. J. (2003). Imitation of successful behaviour in cournot markets. *The Economic Journal*, 113(487):495–524.

Bravo, M. and Mertikopoulos, P. (2017). On the robustness of learning in games with stochastically perturbed payoff observations. *Games and Economic Behavior*, 103:41–66.

Brookins, P. and Ryvkin, D. (2014). An experimental study of bidding in contests of incomplete information. *Experimental Economics*, 17(2):245–261.

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

Camerer, C. F. (2003). *Behavioral game theory*. Russell Sage Foundation New York.

Charness, G., Frechette, G. R., and Kagel, J. H. (2004). How robust is laboratory gift exchange? *Experimental Economics*, 7(2):189–205.

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

Cho, I.-K. and Matsui, A. (2005). Learning aspiration in repeated games. *Journal of Economic Theory*, 124(2):171–201.

Chowdhury, S. M., Mukherjee, A., and Turocy, T. L. (2020). That's the ticket: explicit lottery randomisation and learning in Tullock contests. *Theory and Decision*, 88:405–429.

Chowdhury, S. M., Sheremeta, R. M., and Turocy, T. L. (2014). Overbidding and overspreading in rent-seeking experiments: Cost structure and prize allocation rules. *Games and Economic Behavior*, 87:224–238.

Cornes, R. and Hartley, R. (2003). Risk aversion, heterogeneity and contests. *Public Choice*, 117(1):1–25.

Cornes, R., Hartley, R., et al. (2003). Loss aversion and the Tullock paradox. Working paper, University of Nottingham Discussion Paper No. 03/17.

Cox, C. A. (2017). Rent-seeking and competitive preferences. *Journal of Economic Psychology*, 63:102–116.

Cox, J. C., Smith, V. L., and Walker, J. M. (1988). Theory and individual behavior of first-price auctions. *Journal of Risk and Uncertainty*, 1(1):61–99.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, 47(2):315–72.

Duffy, J. and Hopkins, E. (2005). Learning, information, and sorting in market entry games: theory and evidence. *Games and Economic behavior*, 51(1):31–62.

Engelbrecht-Wiggans, R. and Katok, E. (2009). A direct test of risk aversion and regret in first price sealed-bid auctions. *Decision Analysis*, 6(2):75–86.

Erev, I. and Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological review*, 112(4):912.

Erev, I. and Haruvy, E. (2016). Learning and the economics of small decisions. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics, Volume 2*, chapter 10, pages 638–716. Princeton university press, Princeton, NJ.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881.

Ert, E. and Erev, I. (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behavioral Decision Making*, 20(3):305–322.

Fallucchi, F., Renner, E., and Sefton, M. (2013). Information feedback and contest structure in rent-seeking games. *European Economic Review*, 64:223–240.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Filiz-Ozbay, E. and Ozbay, E. Y. (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review*, 97(4):1407–1418.

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

Fonseca, M. A. (2009). An experimental investigation of asymmetric contests. *International Journal of Industrial Organization*, 27(5):582–591.

Foster, D. P. and Young, H. P. (2006). Regret testing: Learning to play Nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1(3):341–367.

Fudenberg, D. and Peysakhovich, A. (2016). Recency, records, and recaps: Learning and nonequilibrium behavior in a simple decision problem. *ACM Transactions on Economics and Computation (TEAC)*, 4(4):1–18.

Gneezy, U., Rustichini, A., and Vostroknutov, A. (2010). Experience and insight in the race game. *Journal of Economic Behavior & Organization*, 75(2):144–155.

Goeree, J. K. and Holt, C. A. (2001). Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review*, pages 1402–1422.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2002). Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory*, 104(1):247–272.

Gonzalez, R. and Wu, G. (1999). On the shape of the probability weighting function. *Cognitive psychology*, 38(1):129–166.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Grosskopf, B., Bereby-Meyer, Y., and Bazerman, M. (2007). On the robustness of the winner's curse phenomenon. *Theory and Decision*, 63(4):389–418.

Grosskopf, B., Erev, I., and Yechiam, E. (2006). Foregone with the wind: Indirect payoff information and its implications for choice. *International Journal of Game Theory*, 34(2):285–302.

Grund, C. and Sliwka, D. (2005). Envy and compassion in tournaments. *Journal of Economics & Management Strategy*, 14(1):187–207.

Hasselt, H. V. (2010). Double q-learning. In *Advances in Neural Information Processing Systems*, pages 2613–2621.

Herrmann, B. and Orzen, H. (2008). The appearance of homo rivalis: Social preferences and the nature of rent seeking. Working paper, CeDEx discussion paper series.

Hertwig, R., Barron, G., Weber, E. U., and Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological science*, 15(8):534–539.

Hillman, A. L. and Katz, E. (1984). Risk-averse rent seekers and the social cost of monopoly power. *The Economic Journal*, 94(373):104–110.

Ho, T. H., Wang, X., and Camerer, C. F. (2008). Individual differences in ewa learning with partial payoff information. *The Economic Journal*, 118(525):37–59.

Hoffmann, M. and Kolmar, M. (2017). Distributional preferences in probabilistic and share contests. *Journal of Economic Behavior & Organization*, 142:120–139.

Hopkins, E. (2002). Two competing models of how people learn in games. *Econometrica*, 70(6):2141–2166.

Huck, S., Normann, H.-T., and Oechssler, J. (1999). Learning in cournot oligopoly–an experiment. *The Economic Journal*, 109(454):80–95.

Huttegger, S. M. (2013). Probe and adjust. *Biological Theory*, 8(2):195–200.

Jiao, P. and Nax, H. H. (2018). How you learn depends on when you learn: Experimental evidence from cournot contests. *Available at SSRN: https://ssrn.com/abstract=3053958.*

Jindapon, P. and Whaley, C. A. (2015). Risk lovers and the rent over-investment puzzle. *Public Choice*, 164(1-2):87–101.

Jindapon, P. and Yang, Z. (2017). Risk attitudes and heterogeneity in simultaneous and sequential contests. *Journal of Economic Behavior & Organization*, 138:69–84.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kong, X. (2008). Loss aversion and rent-seeking: An experimental study. Working paper, CeDEx discussion paper No. 2008-–13.

Lehrer, E. and Kalai, E. (1993). Rational learning leads to Nash equilibrium. *Econometrica*, 61(5):1019–1045.

Lim, W., Matros, A., and Turocy, T. L. (2014). Bounded rationality and group size in Tullock contests: Experimental evidence. *Journal of Economic Behavior & Organization*, 99:155–167.

Mago, S. D., Samak, A. C., and Sheremeta, R. M. (2016). Facing your opponents: Social identification and information feedback in contests. *Journal of Conflict Resolution*, 60(3):459–481.

Mailath, G. J. (1998). Do people play Nash equilibrium? Lessons from evolutionary game theory. *Journal of Economic Literature*, 36(3):1347–1374.

Masiliūnas, A. (2017). Overcoming coordination failure in a critical mass game: strategic motives and action disclosure. *Journal of Economic Behavior & Organization*, 139:214–251.

Masiliūnas, A., Mengel, F., and Reiss, J. P. (2014). Behavioral variation in Tullock contests. Working paper, Working Paper Series in Economics, Karlsruher Institut für Technologie (KIT).

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Millner, E. L. and Pratt, M. D. (1991). Risk aversion and rent-seeking: An extension and some experimental evidence. *Public Choice*, 69(1):81–92.

Myers, J. L. and Sadler, E. (1960). Effects of range of payoffs as a variable in risk taking. *Journal of Experimental Psychology*, 60(5):306.

Nax, H. H., Burton-Chellew, M. N., West, S. A., and Young, H. P. (2016). Learning in a black box. *Journal of Economic Behavior & Organization*, 127:1–15.

Ockenfels, A. and Selten, R. (2005). Impulse balance equilibrium and feedback in first price auctions. *Games and Economic Behavior*, 51(1):155–170.

Ockenfels, A. and Selten, R. (2014). Impulse balance in the newsvendor game. *Games and Economic Behavior*, 86:237–247.

Otto, A. R. and Love, B. C. (2010). You don't want to know what you're missing: When information about forgone rewards impedes dynamic decision making. *Judgment and Decision Making*, 5(1):1–10.

Parco, J. E., Rapoport, A., and Amaldoss, W. (2005). Two-stage contests with budget constraints: An experimental study. *Journal of Mathematical Psychology*, 49(4):320–338.

Perea, A. (2007). A one-person doxastic characterization of Nash strategies. *Synthese*, 158(2):251–271.

Price, C. R. and Sheremeta, R. M. (2011). Endowment effects in contests. *Economics Letters*, 111(3):217–219.

Price, C. R. and Sheremeta, R. M. (2015). Endowment origin, demographic effects, and individual preferences in contests. *Journal of Economics & Management Strategy*, 24(3):597–619.

Rakow, T., Newell, B. R., and Wright, L. (2015). Forgone but not forgotten: the effects of partial and full feedback in "harsh" and "kind" environments. *Psychonomic bulletin & review*, 22(6):1807–1813.

Riechmann, T. (2007). An analysis of rent-seeking games with relative-payoff maximizers. *Public Choice*, 133(1):147–155.

Roth, A. E. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212.

Salmon, T. C. (2001). An evaluation of econometric models of adaptive learning. *Econometrica*, 69(6):1597–1628.

Savikhin, A. C. and Sheremeta, R. M. (2013). Simultaneous decision-making in competitive and cooperative environments. *Economic Inquiry*, 51(2):1311–1323.

Schmitt, P., Shupp, R., Swope, K., and Cadigan, J. (2004). Multi-period rent-seeking contests with carryover: Theory and experimental evidence. *Economics of Governance*, 5(3):187–211.

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

Selten, R. and Chmura, T. (2008). Stationary concepts for experimental 2x2-games. *American Economic Review*, 98(3):938–66.

Selten, R. and Stoecker, R. (1986). End behavior in sequences of finite prisoner's dilemma supergames a learning theory approach. *Journal of Economic Behavior & Organization*, 7(1):47–70.

Shaffer, S. (2006). Contests with interdependent preferences. *Applied Economics Letters*, 13(13):877–880.

Shafran, A. P. (2012). Learning in games with risky payoffs. *Games and Economic Behavior*, 75(1):354–371.

Sheremeta, R. M. (2010). Experimental comparison of multi-stage and one-stage contests. *Games and Economic Behavior*, 68(2):731–747.

Sheremeta, R. M. (2011). Contest design: An experimental investigation. *Economic Inquiry*, 49(2):573–590.

Sheremeta, R. M. (2013). Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys*, 27(3):491–514.

Sheremeta, R. M. (2016). Impulsive behavior in competition: Testing theories of overbidding in rent-seeking contests. ESI Working Paper 16-21.

Sheremeta, R. M., Masters, W. A., and Cason, T. N. (2018). Winner-take-all and proportional-prize contests: theory and experimental results. *Journal of Economic Behavior & Organization*.

Sheremeta, R. M. and Zhang, J. (2010). Can groups solve the problem of over-bidding in contests? *Social Choice and Welfare*, 35(2):175–197.

Shupp, R., Sheremeta, R. M., Schmidt, D., and Walker, J. (2013). Resource allocation contests: Experimental evidence. *Journal of Economic Psychology*, 39:257–267.

Skaperdas, S. and Gan, L. (1995). Risk aversion in contests. *Economic Journal*, 105(431):951–962.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, 119(1):3.

Thrun, S. and Schwartz, A. (1993). Issues in using function approximation for reinforcement learning. In *Proceedings of the 1993 Connectionist Models Summer School Hillsdale, NJ. Lawrence Erlbaum*.

Tullock, G. (1980). Efficient rent seeking. *Toward a theory of the rent-seeking society*, 97:112.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.

Weber, R. A. (2003). 'Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior*, 44(1):134–144.

Weibull, J. W. (1997). *Evolutionary game theory*. MIT press.

Wilcox, N. T. (2006). Theories of learning in games and heterogeneity bias. *Econometrica*, 74(5):1271–1292.

Wulff, D. U., Mergenthaler-Canesco, M., and Hertwig, R. (2018). A meta-analytic review of two modes of learning and the description-experience gap. *Psychological Bulletin*, 144(2):140–176.

Yechiam, E. and Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, 19(1):1–16.

Yechiam, E. and Rakow, T. (2012). The effect of foregone outcomes on choices from experience. *Experimental psychology*.

Young, H. P. (2009). Learning by trial and error. *Games and Economic Behavior*, 65(2):626–643.

Appendix

A Utility functions

Table 7: A list of possible outcomes in each treatment

Treatment	**Contest**		**Non-contest**		**Payoff**		**Probability**
	i	j	i	j	i	j	
RR	W	L	W	W	$2V$	V	$p_i^c p_i^{nc} p_j^{nc}$
	W	L	W	L	$2V$	0	$p_i^c p_i^{nc}(1-p_j^{nc})$
	W	L	L	W	V	V	$p_i^c(1-p_i^{nc})p_j^{nc}$
	W	L	L	L	V	0	$p_i^c(1-p_i^{nc})(1-p_j^{nc})$
	L	L	W	W	V	$2V$	$(1-p_i^c)p_i^{nc}p_j^{nc}$
	L	L	W	L	V	V	$(1-p_i^c)p_i^{nc}(1-p_j^{nc})$
	L	L	L	W	0	$2V$	$(1-p_i^c)(1-p_i^{nc})p_j^{nc}$
	L	L	L	L	0	V	$(1-p_i^c)(1-p_i^{nc})(1-p_j^{nc})$
SR	W	L	–	–	$V+E-c_i$	$E-c_j$	p_i^c
	L	W	–	–	$E-c_i$	$V+E-c_j$	$1-p_i^c$
RS	–	–	W	W	$V+\frac{c_i}{c_i+c_j}V$	$V+\frac{c_j}{c_i+c_j}V$	$p_i^{nc}p_j^{nc}$
	–	–	W	L	$V+\frac{c_i}{c_i+c_j}V$	$\frac{c_j}{c_i+c_j}V$	$p_i^{nc}(1-p_j^{nc})$
	–	–	L	W	$\frac{c_i}{c_i+c_j}V$	$V+\frac{c_j}{c_i+c_j}V$	$(1-p_i^{nc})p_j^{nc}$
	–	–	L	L	$\frac{c_i}{c_i+c_j}V$	$\frac{c_j}{c_i+c_j}V$	$(1-p_i^{nc})(1-p_j^{nc})$
SS	–	–	–	–	$E-c_i+\frac{c_i}{c_i+c_j}V$	$E-c_j+\frac{c_j}{c_i+c_j}V$	1

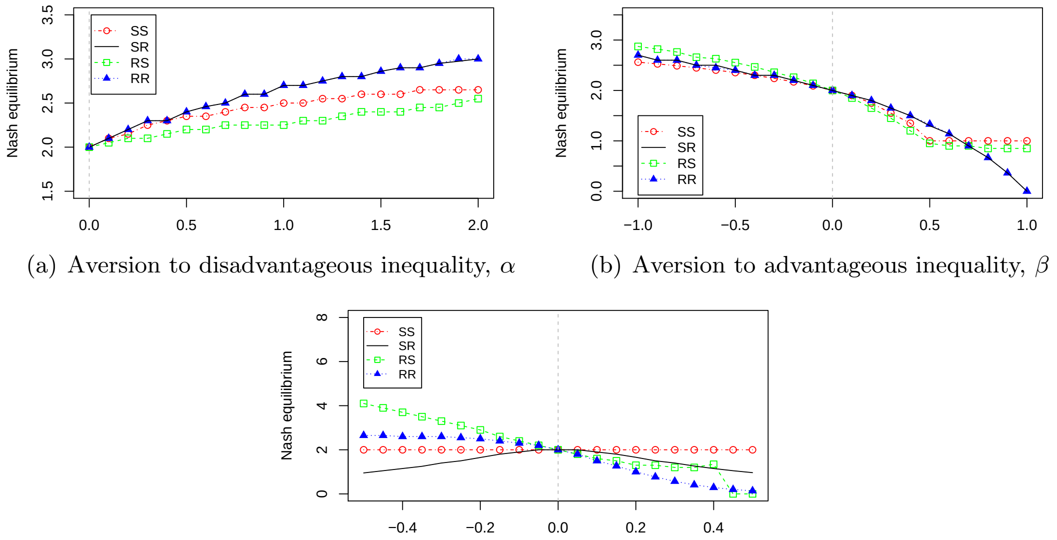
Table 7 lists all possible outcomes in each treatment. For each outcome, we specify payoffs obtained by player i and opponent j, whether each player won (W) or lost (L) the reward from the contest and non-contest lotteries and the probability of the outcome.

The calculation of utility requires additional assumptions about how social preferences, non-monetary utility of winning and probability weighting are interpreted in contests. For social preferences, players could care about the distribution of expected payoffs (Hoffmann and Kolmar, 2017) or about realized payoffs (Herrmann and Orzen, 2008, Fonseca, 2009). Some studies assume preferences over the distribution of total earnings (Herrmann and Orzen, 2008, Fonseca, 2009, Mago et al., 2016 Shaffer, 2006), while others assume preferences only over the distribution of lottery earnings, net of investments costs (Grund and Sliwka, 2005, Hoffmann and Kolmar, 2017). We assume preferences over ex-post outcomes, including earnings from all sources. If we instead assumed preferences over ex-ante payoffs, predictions in all treatments would coincide with those in SS. Finally, we could represent social preferences using a model of spitefulness or inequality aversion. We choose the spitefulness specification because it is simpler to model and has a single parameter, and it can explain a wider range of choices, therefore we expect it to fit data better. But we also calculate equilibria under inequality aversion and show that the predictions are quite similar.

For non-monetary utility from winning, in section 4.3 we assume that players receive the same utility from winning the contest prize as from winning the non-contest prize. In section 5.2 we allow non-monetary utility to also depend on the type of lottery won.

Probability weighting is straightforward in SS (because there is only one outcome) and in SR (because there are two outcomes). In SR, the outcome in which the player receives the reward always provides a higher payoff, therefore it is weighted by $w(p_i^c)$, while the other outcome is weighted by $(1-w(p_i^c))$. In RS, two pairs of outcomes are possible, but each pair provides identical payoffs for i. Cumulative prospect theory requires a strict order of outcomes, therefore all outcomes with identical outcomes must be combined. Unfortunately, outcomes cannot be easily combined because some outcomes will have identical payoffs but different utilities, because of social preferences. We therefore separately calculate the probability that each pair of outcomes will occur (p_i^{nc} and $(1-p_i^{nc})$), weight these probabilities using $w(\cdot)$ and then divide the weights between outcomes in each pair proportionally to their (unweighted) probabilities. We could have weighted the probabilities within each pair too, but chose not to as it is unclear how probability weighting operates with the payoffs of the other participant. Our approach ensures that players with no social preferences always weight the outcomes exactly as in the cumulative prospect theory. We follow the same approach in RR, in which there are eight possible outcomes but only three distinct payoffs for i. We first calculate probabilities for each group of outcomes that generate these three payoff levels, and weight them. The outcome group with payoff $2V$ receives weight $w(p_i^c p_i^{nc})$. The outcome group with a payoff of V receives weight $w(p(2V)+p(V)) - w(p(2V)) = w(p_i^c p_i^{nc} + p_i^c(1-p_i^{nc}) + (1-p_i^c)p_i^{nc})) - w(p_i^c p_i^{nc})$. The outcome group with the payoff of 0 receives the remaining weight, ensuring that probabilities add up to 1. These weights are then divided within each outcome group proportionally to the (unweighted) probabilities of each individual outcome.

B Alternative utility specifications



(a) Aversion to disadvantageous inequality, α

(b) Aversion to advantageous inequality, β

(c) CARA, a

Figure 9: Nash equilibrium with different model specifications.

- **Constant absolute risk aversion (CARA).** We use an exponential function, in which utility equals payoff under risk neutrality ($a = 0$), while risk seeking ($a < 0$) and risk averse preferences ($a > 0$) are modeled by:

$$u_i(\pi_i, \pi_j, W_i) = (1 - e^{-a\pi_i})/a \tag{10}$$

- **Inequality aversion.** We follow the Fehr and Schmidt (1999) model, in which the weights of disadvantageous and advantageous inequality are α and β:

$$u_i(\pi_i, \pi_j, W_i) = \begin{cases} \pi_i - \alpha(\pi_j - \pi_i) & \text{if } \pi_i < \pi_j \\ \pi_i - \beta(\pi_i - \pi_j) & \text{if } \pi_i > \pi_j \end{cases}$$

C Impulse balance equilibrium

This section shows the calculation of the impulse balance equilibrium (IBE). Weighted IBE distinguishes between losses and gains, but since participants cannot lose money in our experiments, we need to make an assumption about the reference income level. We assume that the reference point is the initial endowment,[28] which also coincides with the maximin payoff, used as a reference point in IBE (e.g. Selten and Chmura, 2008). Whenever the realized payoff falls below the endowment, the strength of an impulse is multiplied by θ. In line with the prospect theory (Kahneman and Tversky, 1979), it is often assumed that $\theta = 2$ (Selten and Chmura, 2008), but we will study a wider range of values to distinguish between loss aversion and ex-post rationality. For simplicity, we assume that $E = V$, that is the endowment is equal to the value of the prize, as in our experiment and in most other studies.

For each action profile $\{c_i, c_j\}$, with $c_i, c_j \in [0, V]$, we calculate the expected upward ($E^+(c_i, c_j)$) and downward ($E^-(c_i, c_j)$) impulses. The symmetric weighted impulse balance equilibrium is an action profile $\{c^*, c^*\}$ in which upward and downward impulses are equalized, that is $E^+(c^*, c^*) = E^-(c^*, c^*)$.

The ex-post rational action will be determined by the lottery outcome in SR and RS, and by the outcome of two lotteries in RR. There are no IBE in which both players invest nothing into the contest because there would be no downward impulse and a positive expected upward impulse, in all treatments. We can therefore calculate the probability to win the contest by $p_c = \frac{c_i}{c_i + c_j}$.

C.1 SR

The probability that i receives the prize in profile $\{c_i, c_j\}$ is $p_c = \frac{c_i}{c_i + c_j}$. The prize is won if a random variable $r_c \in \mathcal{U}(0, 1)$ exceeds p_c. In our experimental design, r_c was generated once per round, therefore if a prize was won, it would have also been won with a higher contest investment level, and if the prize was not won, it would not have been won with a lower investment level. If the win is possible, it is ex-post optimal to choose an action such that the probability to win is exactly equal to the generated number: $r_c = \frac{c_i^*}{c_i^* + c_j}$, therefore the ex-post rational action is $c_i^* = \frac{r_c c_j}{1 - r_c}$. If the win is not possible ($r_c > \frac{V}{V + c_j}$), the ex-post rational contest investment is 0.

Table 8 shows that there are three possible outcomes. In the first case, player suffers from wasted investment: the prize was won, but it could have been won by a lower investment level. If the second case, the player suffers from lost opportunity: the prize was not won,

[28] The assumption that the reference point is equal to the endowment has been implicitly made in papers that study loss aversion in contests (Cornes et al., 2003; Kong, 2008).

Outcome	Condition	$\pi(c_i, c_j)$	$\pi(c_i^*, c_j)$	Impulse size	Direction	Loss
Wasted investment	$r_c \in (0, p_c)$	$V - c_i + V$	$V - \frac{r_c c_j}{1-r_c} + V$	$c_i - \frac{r_c c_j}{1-r_c}$	↓	No
Lost opportunity	$r_c \in (p_c, \frac{V}{V+c_j}]$	$V - c_i$	$V - \frac{r_c c_j}{1-r_c} + V$	$c_i - \frac{r_c c_j}{1-r_c} + V$	↑	Yes
Impossible win	$r_c \in (\frac{V}{V+c_j}, 1)$	$V - c_i$	V	c_i	↓	Yes

Table 8: Impulses in SR.

but it could have been won by a sufficiently higher contest investment. In the third case, winning was not possible because of a high generated number, therefore the ex-post optimal investment level is 0. In the latter two cases, players do not win the prize despite a positive investment, therefore a loss is incurred, and impulses are multiplied by θ. We calculate the expected upward and downward impulses by integrating over all possible realizations of r_c.

$$\begin{aligned} E^-(c_i, c_j) &= \int_0^{p_c} c_i - \frac{r_c c_j}{1 - r_c} \,\mathrm{d}r_c + \theta \int_{\frac{V}{V+c_j}}^1 c_i \,\mathrm{d}r_c = \\ &= c_i + c_j \log\left(\frac{c_j}{c_i + c_j}\right) + \frac{\theta c_i c_j}{V + c_j} \end{aligned}$$

$$\begin{aligned} E^+(c_i, c_j) &= \theta \int_{p_c}^{\frac{V}{V+c_j}} c_i - \frac{r_c c_j}{1 - r_c} + V \,\mathrm{d}r_c = \\ &= \frac{(V + c_i + c_j)(V - c_i)\theta c_j}{(V + c_j)(c_i + c_j)} + \theta c_j \log\left(\frac{c_i + c_j}{V + c_j}\right) \end{aligned}$$

Action profile $\{c^*, c^*\}$ is a symmetric weighted impulse balance equilibrium if it satisfies $E^-(c^*, c^*) = E^+(c^*, c^*)$. Simplifying and rearranging gives the following condition:

$$\frac{\theta V}{2c^*} + \frac{2\theta V}{V + c^*} - 2\theta - 1 - \log(0.5) + \theta \log\left(\frac{2c^*}{V + c^*}\right) = 0 \tag{11}$$

For the parameters used in the experiment ($V = 8$), IBE predicts over-investments even without loss aversion ($c^* = 2.66$ if $\theta = 1$), and investments are increasing in loss aversion ($c^* = 3.09$ if $\theta = 2$).

C.2 RS

In RS, the probability that player i receives the prize is $p_{nc} = \frac{V - c_i}{V}$. The prize is received if $r_{nc} \leqslant p_{nc}$, where r_{nc} is drawn uniformly from $[0, 1]$. Then the range of p_{nc} is $[0, 1]$, and the ex-post rational action must satisfy $p_{nc} = r_{nc}$, therefore $c_i^* = V - r_{nc}V$.

Table 9 shows that two types of impulses are possible. If $r_{nc} < p_{nc}$, the player suffers from a wasted investment: the non-contest prize was won, but it could have been won by investing more into the contest (upward impulse). If $r_{nc} > p_{nc}$, the player suffers from a lost opportunity: the prize was not won, but it could have been won if the contest investment was sufficiently reduced (downward impulse). In the case of a lost opportunity, earnings are below the initial endowment, therefore the downward impulse is multiplied by θ. Expected impulses are calculated by integrating over the possible values of r_{nc}:

Outcome	Condition	$\pi(c_i, c_j)$	$\pi(c_i^*, c_j)$	Impulse size	Direction	Loss
Wasted investment	$r_{nc} \in (0, p_{nc})$	$\frac{c_i}{c_i+c_j}V + V$	$\frac{V-r_{nc}V}{V-r_{nc}V+c_j}V + V$	$(\frac{V-r_{nc}V}{V-r_{nc}V+c_j} - \frac{c_i}{c_i+c_j})V$	↑	No
Lost opportunity	$r_{nc} \in (p_{nc}, 1)$	$\frac{c_i}{c_i+c_j}V$	$\frac{V-r_{nc}V}{V-r_{nc}V+c_j}V + V$	$(\frac{V-r_{nc}V}{V-r_{nc}V+c_j} + \frac{c_j}{c_i+c_j})V$	↓	Yes

Table 9: Impulses in RS.

$$\begin{aligned}E^+(c_i, c_j) &= \int_0^{p_{nc}} \left(\frac{V - r_{nc}V}{V - r_{nc}V + c_j} - \frac{c_i}{c_i + c_j} \right) V \, \mathrm{d}r_{nc} = \\ &= (V - c_i)\frac{c_j}{c_i + c_j} + c_j \log\left(\frac{c_i + c_j}{V + c_j} \right)\end{aligned}$$

$$\begin{aligned}E^-(c_i, c_j) &= \theta \int_{p_{nc}}^1 \left(\frac{V - r_{nc}V}{V - r_{nc}V + c_j} + \frac{c_j}{c_i + c_j} \right) V \, \mathrm{d}r_{nc} = \\ &= \frac{\theta c_i(c_i + 2c_j)}{c_i + c_j} + \theta c_j \log\left(\frac{c_j}{c_i + c_j} \right)\end{aligned}$$

Setting $E^-(c^*, c^*) = E^+(c^*, c^*)$ and simplifying gives the following condition:

$$c^*(0.5 + (1.5 + \log(0.5))\theta) - c^* \log\left(\frac{2c^*}{V + c^*} \right) - 0.5V = 0 \tag{12}$$

If $V = 8$ and $\theta = 1$, $c^* = 1.7$. Loss aversion reduces contest investment even further, because it increases the downward impulse.

C.3 RR

In RR, players face a tradeoff between the probability to win the contest prize ($p_c = \frac{c_i}{c_i+c_j}$) and the probability to win the non-contest prize ($p_{nc} = \frac{V-c_i}{V}$). Figure 10 illustrates this tradeoff using the "budget line" (plotted in blue), calculated as $p_c = \frac{V(1-p_{nc})}{V(1-p_{nc})+c_j}$. If nothing is invested into the contest, $p_{nc} = 1$ and $p_c = 0$ (bottom right corner); if the entire endowment is invested into contest, $p_{nc} = 0$ and $p_c = \frac{V}{V+c_j}$ (top left corner). Lottery outcomes are determined by numbers $r_c, r_{nc} \in \mathcal{U}(0, 1)$. Lottery outcome can be represented by a point on a unit square, and the corresponding prize is won if the generated number is closer to the origin than the probability to win.

Figure 10 illustrates impulses for the RNNE action, marked on the line. If the lottery outcome falls in area a, both prizes are won, therefore the chosen action maximizes ex-post payoffs. If the lottery outcome is in area b, the player receives only the non-contest prize, and there is a higher contest investment level that would have resulted in winning both prizes, therefore an upward impulse of size V is received. If the lottery outcome is in area c, only the contest prize is received, while both prizes could have been won by choosing a lower contest investment level, therefore a downward impulse of size V is received. Only the non-contest prize is received in areas d and e, and only the contest prize is received in area f; these areas lie outside of the budget set, therefore both prizes could not have been won, and no impulse is received. In area g, no prize is received, but a prize could have been won by either a lower or a higher contest investment. We assume that no impulse is received in that case. In area h,

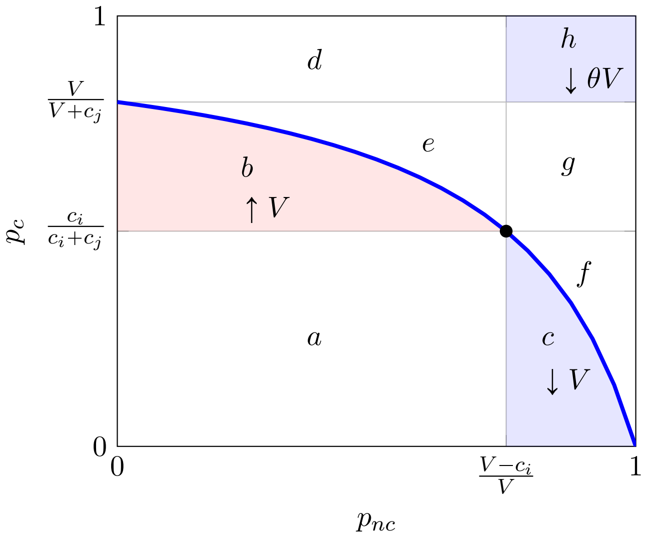


Figure 10: Impulses in RR. The marked location and impulses are plotted for the RNNE action ($c_i = V/4$). The "budget line", marked in blue, shows the boundary of all attainable combinations of probabilities, if the opponent chooses the RNNE action ($c_j = V/4$).

no prize is received, but a non-contest prize could have been received by a sufficiently lower contest investment, therefore a downward impulse is received. If the lottery outcome falls in area h, payoffs fall below the endowment, therefore the impulse is multiplied by the loss aversion parameter θ.

Players receive an upward impulse if the outcome is in area b, and a downward impulse if it is in c or h. We calculate these probabilities by integrating the area under the budget line:

$$b = \int_0^{\frac{V-c_i}{V}} \frac{V(1-p_{nc})}{V(1-p_{nc})+c_j} - \frac{c_i}{c_i+c_j} \, \mathrm{d}p_{nc} = \frac{c_j}{V} \log\left(\frac{c_i+c_j}{V+c_j}\right) + \frac{c_j(V-c_i)}{V(c_i+c_j)}$$

$$c = \int_{\frac{V-c_i}{V}}^{1} \frac{V(1-p_{nc})}{V(1-p_{nc})+c_j} \, \mathrm{d}p_{nc} = \frac{c_j}{V} \log\left(\frac{c_j}{c_i+c_j}\right) + \frac{c_i}{V}$$

$$h = \left(1 - \frac{V-c_i}{V}\right)\left(1 - \frac{V}{V+c_j}\right) = \frac{c_i c_j}{V(V+c_j)}$$

Expected impulse is a product of impulse probability and strength: $E^+(c_i, c_j) = bV$, $E^-(c_i, c_j) = cV + h\theta V$. A symmetric impulse balance equilibrium $\{c^*, c^*\}$ must equalize upward and downward impulses. Substituting the definition of b, c and h and simplifying gives the following condition that c^* must satisfy:

$$\log\left(\frac{4c^*}{V+c^*}\right) - \frac{3c^*-V}{2c^*} - \theta\frac{c^*}{V+c^*} = 0 \tag{13}$$

If $V = 8$ and $\theta = 1$, $c^* = 2.12$, but increased loss aversion increases the downward impulse, therefore IBE falls below RNNE when loss aversion is sufficiently strong.

In SS, the ex-post rational action is always equal to the best-response, thus no player receives an impulse only if the actions of both players are mutual best-responses. Therefore, IBE in the SS treatment coincides with the RNNE.

Each of the conditions (11), (12) and (13) are satisfied by a unique c^*, which depends on the loss aversion parameter θ. These IBE predictions are compared in panel (a) of figure 4. Without loss aversion, IBE is above RNNE in SR, but below it in RS. This difference is a result of a string impulse after failing to win the prize. Expectation of this impulse induces players to over-invest into activity that can potentially provide the prize. Loss aversion further increases the intensity of lost opportunity, and therefore increases the gap between SR and RS. RR treatment falls between SR and RS, because ex-post rational actions can lie on either side of the chosen action.

D Supplementary results

Table 10: Mann-Whitney U test two-sided p-values, for a pairwise comparison of contest investments, averaged by matching group (6 independent observations per treatment).

Rounds 1–10				Rounds 31–40			
Treatment	SR	RS	SS	Treatment	SR	RS	SS
RR	0.75	0.52	0.11	RR	0.47	0.34	0.016
SR	–	0.20	0.01	SR	–	0.078	0.0039
RS	–	–	0.30	RS	–	–	0.0039

Table 11: Mann-Whitney U test two-sided p-values, for a comparison of standard deviation of choices, calculated for each matching group (6 independent observations per treatment).

Rounds 1–10				Rounds 31–40			
Treatment	SR	RS	SS	Treatment	SR	RS	SS
RR	0.52	0.34	0.87	RR	0.20	0.63	0.0039
SR	–	0.34	0.63	SR	–	0.75	0.0039
RS	–	–	0.75	RS	–	–	0.0039

Table 12: Mann-Whitney U test two-sided p-values, for a pairwise comparison of RNNE frequency, averaged by matching group (6 independent observations per treatment).

Rounds 1–10				Rounds 31–40			
Treatment	SR	RS	SS	Treatment	SR	RS	SS
RR	0.73	0.13	0.20	RR	0.36	0.18	0.0038
SR	–	0.03	0.31	SR	–	0.016	0.0036
RS	–	–	0.81	RS	–	–	0.0033

Table 13: Mann-Whitney U test two-sided p-values, for a pairwise comparison of dominated strategy frequency, averaged by matching group (6 independent observations per treatment).

Rounds 1–10				Rounds 31–40			
Treatment	SR	RS	SS	Treatment	SR	RS	SS
RR	0.42	0.42	0.11	RR	0.20	0.13	0.0038
SR	–	0.13	0.01	SR	–	0.024	0.0038
RS	–	–	0.69	RS	–	–	0.0038

E Comparison to other contest experiments

This section tests whether the improvement in the explanatory power observed in SS when FPI was introduced is different from the typical learning rate in proportional contests. We compare our SS treatment to three earlier studies that investigated proportional contests. We thank the authors of these papers for sharing the data. The first dataset is from Masiliūnas et al. (2014), who use the same subject pool (students at Maastricht University), similar framing and experimental design (a two-player game with random matching within a 6-person matching group). As in our paper, the contest was repeated 40 times, but the first 5 rounds of each 10-round block were not incentivized, and are therefore excluded. The prize value and endowment were equal to 16. The second dataset is from the SHARE-FULL treatment in Fallucchi et al. (2013), in which 3 players compete in fixed groups for a prize of 1000 points for 60 rounds. The third dataset is from SL treatment in Chowdhury et al. (2014), with groups of 4 players, random matching and a prize of 80 points, repeatad for 30 rounds. Since each experiment used a different prize value and strategy space, the fraction of RNNE choices would

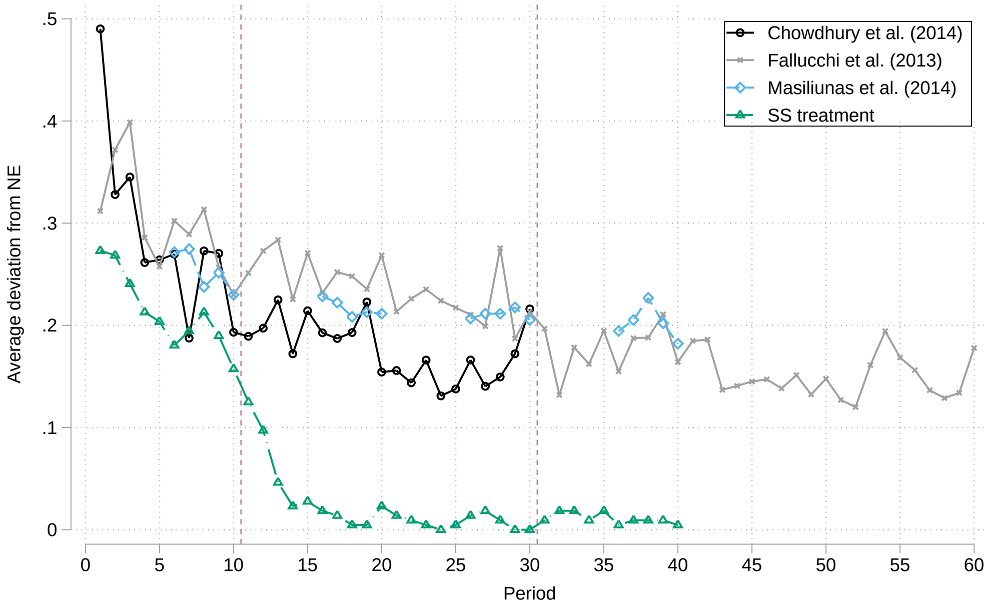


Figure 11: Deviations from RNNE in proportional contests from Chowdhury et al. (2014), Fallucchi et al. (2013), Masiliūnas et al. (2014) and SS treatment.

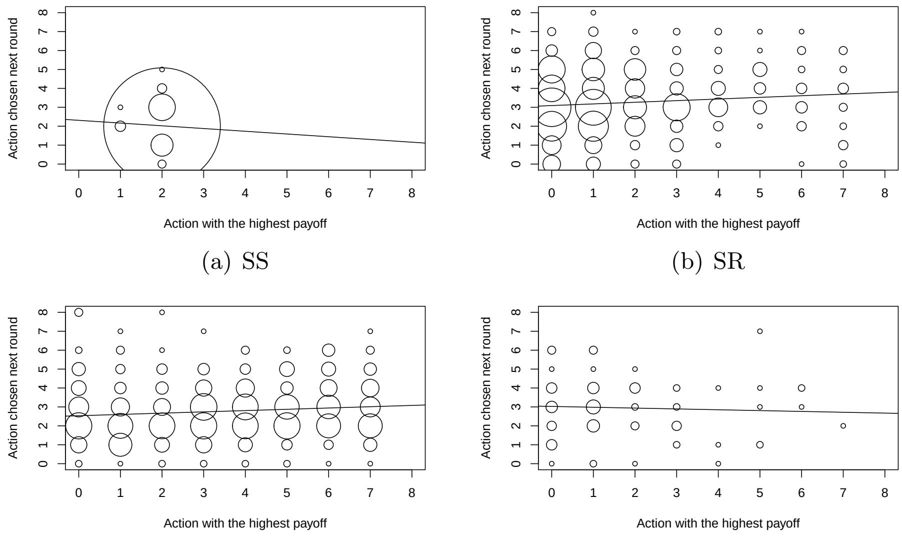
not be comparable. To facilitate comparability we calculated the average absolute value of the difference between choice and RNNE, normalized between 0 and 1 (0 if everyone chose RNNE, 1 if everyone made the largest possible deviation, i.e. invested the entire endowment). Figure 11 plots this measure of equilibrium deviations. Despite the differences in design and in the subject pool, the other three studies find a similar rate of equilibrium deviations. Equilibrium deviations in SS are somewhat below the other three studies at the start of the game, perhaps because of a lower cardinality of the strategy space. We evaluate the differences statistically by averaging the equilibrium deviations over the first 10 rounds, for each matching group.[29] The difference between SS and the other three experiments is not statistically significant if treatments are compared pairwise of if data from all three datasets is pooled (lowest Mann-Whitney U test two-sided p-value is 0.103). We repeat the same exercise for rounds 20-30, which are the last 10 rounds in Chowdhury et al. (2014). The difference between SS and the other three studies is significant when other studies are pooled ($p = 0.0002$) or compared pairwise (p=0.0011 compared to Fallucchi et al., 2013, p=0.0014 compared to Masiliūnas et al., 2014, p=0.0196 compared to Chowdhury et al., 2014). The lower significance in the comparison to Chowdhury et al. (2014) is likely a result of there being only three independent observations. We conclude that the rate of equilibrium deviations in SS is not significantly different from standard proportional contests at the start of the game, but significantly lower once FPI has been introduced. All studies find a decrease in equilibrium deviations, but the magnitude is much lower than the decrease in SS.

F Observed feedback and subsequent adaption

Section 5.1 provided evidence that players choose actions that would have maximized the payoff in the previous round, when this information is available. We will explore this result in more detail by comparing treatments in two aspects: (i) which actions are observed maximizing ex-post payoffs in rounds with FPI, and (ii) how this feedback affects subsequent decisions. We will jointly study both aspects to understand whether treatment differences are caused by differences in the quality of feedback or in the adaptation process.

In terms of the adaptation rules, we find that players choose the action that maximized ex-post payoffs, but only when these payoffs are observed. We calculate the action which would have maximized ex-post payoffs in the previous round using data from rounds in which FPI was available and there was a single payoff-maximizing action. Figure 12 displays the joint distributions of ex-post payoff-maximizing investment levels and the investments chosen the following period (area of a bubble is proportional the number of observations). There is little evidence that players choose actions that have performed best in the previous round, and the slope of a regression line (displayed in the graph) is not significantly different from zero, in all treatments. Failure to choose the foregone payoff-maximizing action may occur if players do not observe the relevant FPI. In the experiment, FPI was initially hidden, and had to be uncovered with a mouse click. Information acquisition was costless, but few players acquired information about all actions. Since information acquisition was tracked, it is possible to identify the action that provided the highest payoff from the set of actions whose foregone payoffs were observed. Figure 13 shows that players do choose actions that were observed

[29] From Masiliūnas et al. (2014) we use only rounds 6-10, which were incentivized. Overall, we have 6 independent observations from our SS treatment, 3 from Chowdhury et al. (2014), 9 from Masiliūnas et al. (2014) and 10 from Fallucchi et al. (2013).



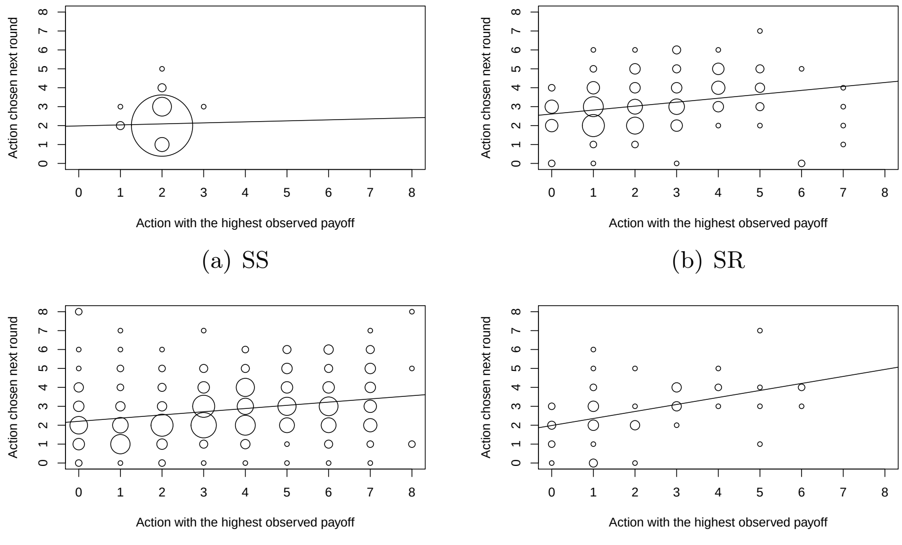
(a) SS

(b) SR

(c) RS

(d) RR

Figure 12: Ex-post payoff-maximizing action, and the action chosen next period. Only data from rounds with FPI in which one action maximizes ex-post payoffs. Area of a bubble is proportional to the number of observations. Regression lines are added to all graphs.



(a) SS

(b) SR

(c) RS

(d) RR

Figure 13: Observed ex-post payoff-maximizing action, and the action chosen next period. Only data from rounds with FPI in which at least one foregone payoff was observed and one action maximized ex-post payoffs. Area of the bubble is proportional to the number of observations. Regression lines are added to all graphs.

providing high payoffs, and the slope of the regression line is significantly higher than zero in SS and RR (p-values respectively 0.055 and 0.026) and marginally significant in SR ($p = 0.11$).

G Figures

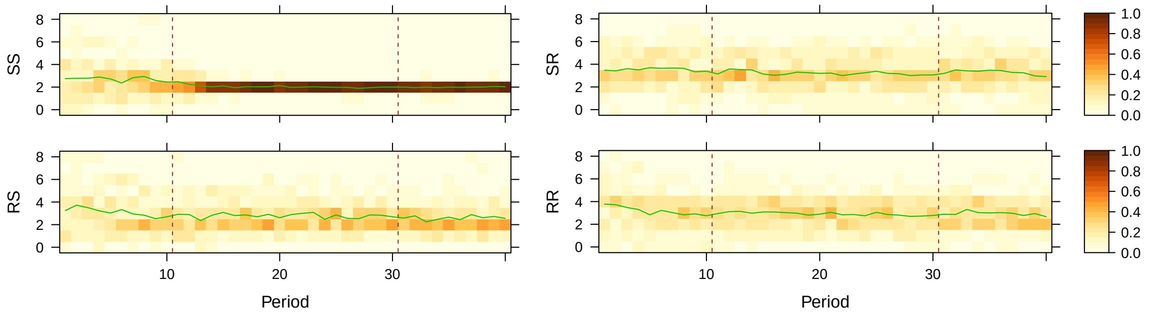
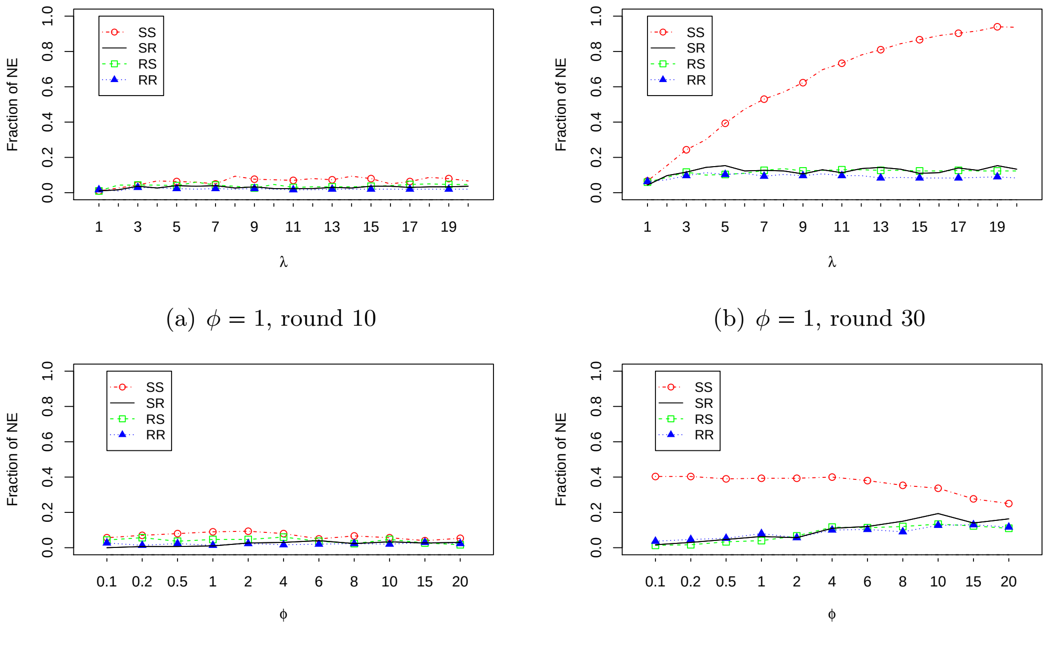


Figure 14: Density of contest investments and their average (green line) in each treatment. FPI was available from round 11 to round 30.



(a) $\phi = 1$, round 10

(b) $\phi = 1$, round 30

(c) $\lambda = 10$, round 10

(d) $\lambda = 10$, round 30

Figure 15: Fraction of RNNE pairs in reinforcement learning simulations. Data only from round 10 and round 30.

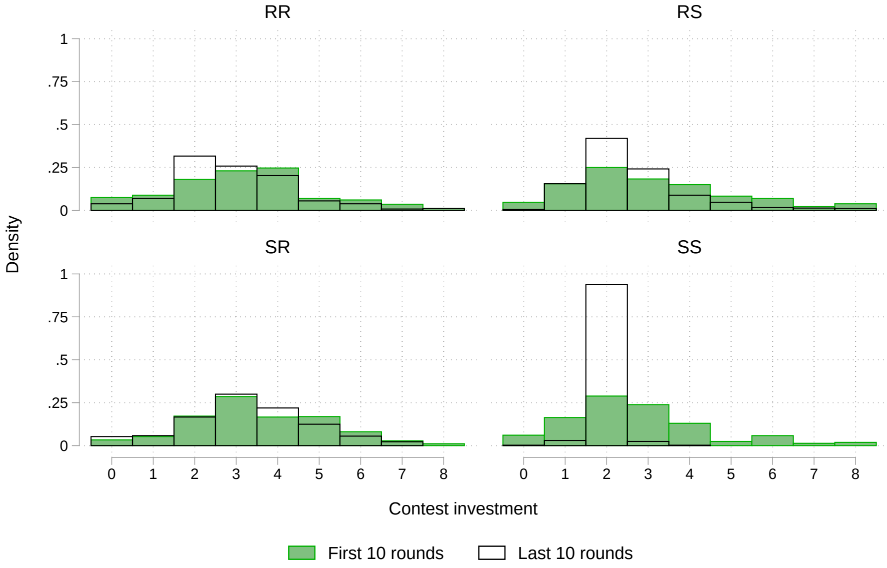


Figure 16: Distribution of choices in the first 10 and in the last 10 rounds, by treatment.

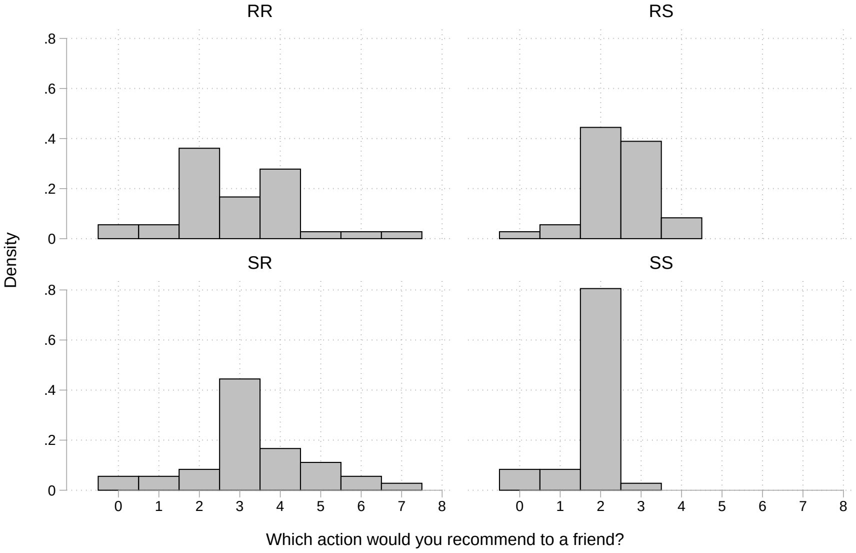
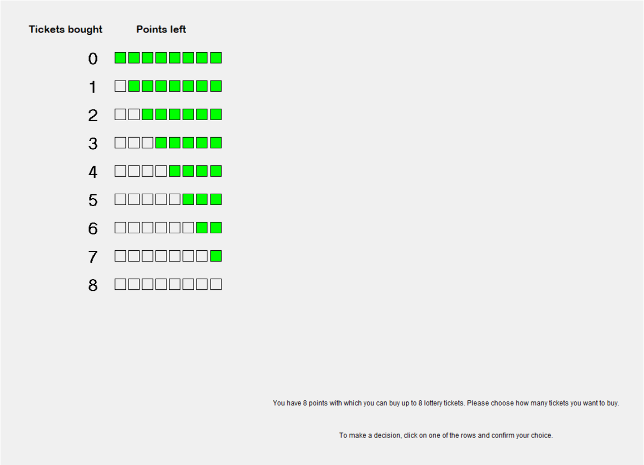


Figure 17: Distribution of actions that players would recommend a friend to play.

H Screenshots

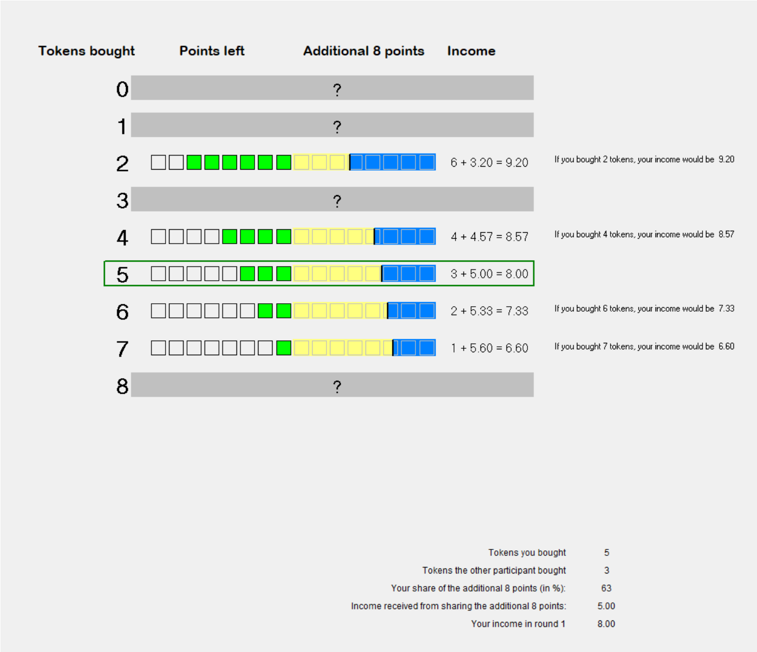


Tickets bought	Points left
0	🟩🟩🟩🟩🟩🟩🟩🟩
1	□🟩🟩🟩🟩🟩🟩🟩
2	□□🟩🟩🟩🟩🟩🟩
3	□□□🟩🟩🟩🟩🟩
4	□□□□🟩🟩🟩🟩
5	□□□□□🟩🟩🟩
6	□□□□□□🟩🟩
7	□□□□□□□🟩
8	□□□□□□□□

You have 8 points with which you can buy up to 8 lottery tickets. Please choose how many tickets you want to buy.

To make a decision, click on one of the rows and confirm your choice.

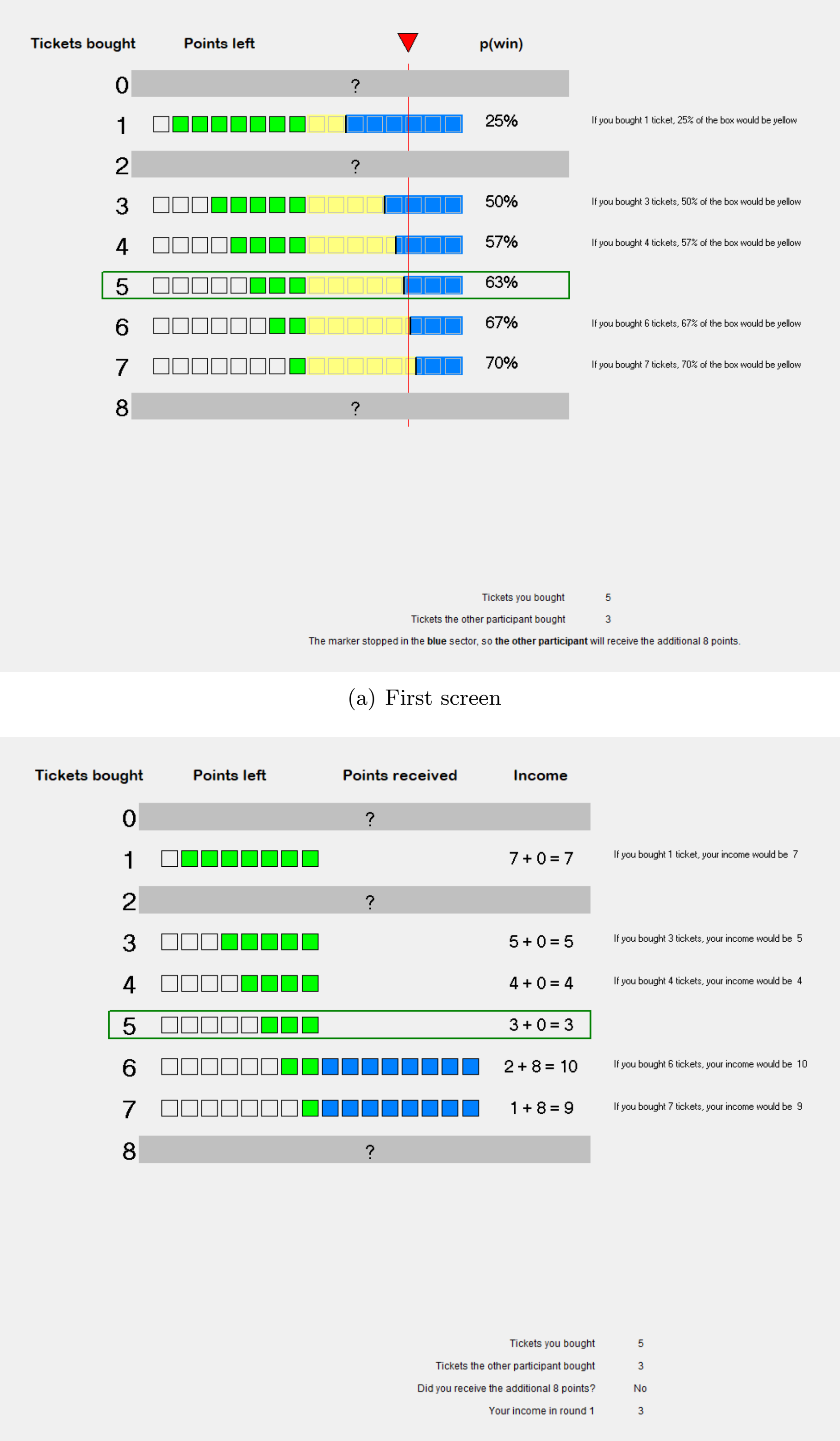
Figure 18: Decision screen in SR treatment.



Tokens bought	Points left	Additional 8 points	Income	
0	?			
1	?			
2	□□🟩🟩🟩🟩🟩🟩		6 + 3.20 = 9.20	If you bought 2 tokens, your income would be 9.20
3	?			
4	□□□□🟩🟩🟩🟩		4 + 4.57 = 8.57	If you bought 4 tokens, your income would be 8.57
5	□□□□□🟩🟩🟩		3 + 5.00 = 8.00	
6	□□□□□□🟩🟩		2 + 5.33 = 7.33	If you bought 6 tokens, your income would be 7.33
7	□□□□□□□🟩		1 + 5.60 = 6.60	If you bought 7 tokens, your income would be 6.60
8	?			

Tokens you bought	5
Tokens the other participant bought	3
Your share of the additional 8 points (in %):	63
Income received from sharing the additional 8 points:	5.00
Your income in round 1	8.00

Figure 19: Feedback in SS treatment with FPI.



Tickets bought	Points left	p(win)	
0	?		
1	□🟩🟩🟩🟩🟩🟩🟩	25%	If you bought 1 ticket, 25% of the box would be yellow
2	?		
3	□□□🟩🟩🟩🟩🟩	50%	If you bought 3 tickets, 50% of the box would be yellow
4	□□□□🟩🟩🟩🟩	57%	If you bought 4 tickets, 57% of the box would be yellow
5	□□□□□🟩🟩🟩	63%	
6	□□□□□□🟩🟩	67%	If you bought 6 tickets, 67% of the box would be yellow
7	□□□□□□□🟩	70%	If you bought 7 tickets, 70% of the box would be yellow
8	?		

Tickets you bought 5

Tickets the other participant bought 3

The marker stopped in the **blue** sector, so **the other participant** will receive the additional 8 points.

(a) First screen

Tickets bought	Points left	Points received	Income	
0	?			
1	□🟩🟩🟩🟩🟩🟩🟩		7 + 0 = 7	If you bought 1 ticket, your income would be 7
2	?			
3	□□□🟩🟩🟩🟩🟩		5 + 0 = 5	If you bought 3 tickets, your income would be 5
4	□□□□🟩🟩🟩🟩		4 + 0 = 4	If you bought 4 tickets, your income would be 4
5	□□□□□🟩🟩🟩		3 + 0 = 3	
6	□□□□□□🟩🟩	🟦🟦🟦🟦🟦🟦🟦🟦	2 + 8 = 10	If you bought 6 tickets, your income would be 10
7	□□□□□□□🟩	🟦🟦🟦🟦🟦🟦🟦🟦	1 + 8 = 9	If you bought 7 tickets, your income would be 9
8	?			

Tickets you bought 5

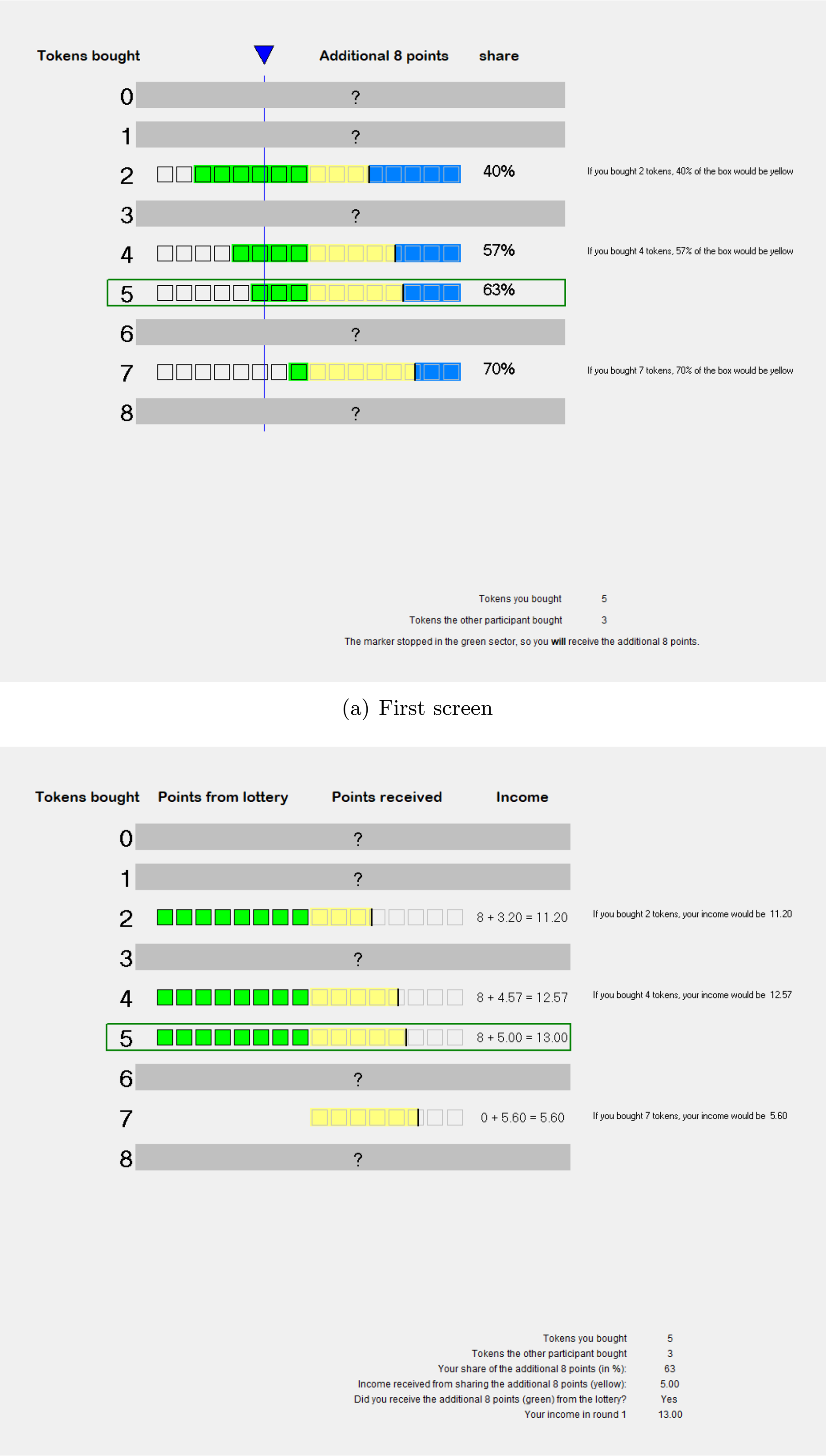
Tickets the other participant bought 3

Did you receive the additional 8 points? No

Your income in round 1 3

(b) Second screen

Figure 20: Feedback in SR treatment with FPI.



Tokens bought | Additional 8 points | share

0 ?

1 ?

2 40% If you bought 2 tokens, 40% of the box would be yellow

3 ?

4 57% If you bought 4 tokens, 57% of the box would be yellow

5 63%

6 ?

7 70% If you bought 7 tokens, 70% of the box would be yellow

8 ?

Tokens you bought 5

Tokens the other participant bought 3

The marker stopped in the green sector, so you **will** receive the additional 8 points.

(a) First screen

Tokens bought | Points from lottery | Points received | Income

0 ?

1 ?

2 8 + 3.20 = 11.20 If you bought 2 tokens, your income would be 11.20

3 ?

4 8 + 4.57 = 12.57 If you bought 4 tokens, your income would be 12.57

5 8 + 5.00 = 13.00

6 ?

7 0 + 5.60 = 5.60 If you bought 7 tokens, your income would be 5.60

8 ?

Tokens you bought 5

Tokens the other participant bought 3

Your share of the additional 8 points (in %): 63

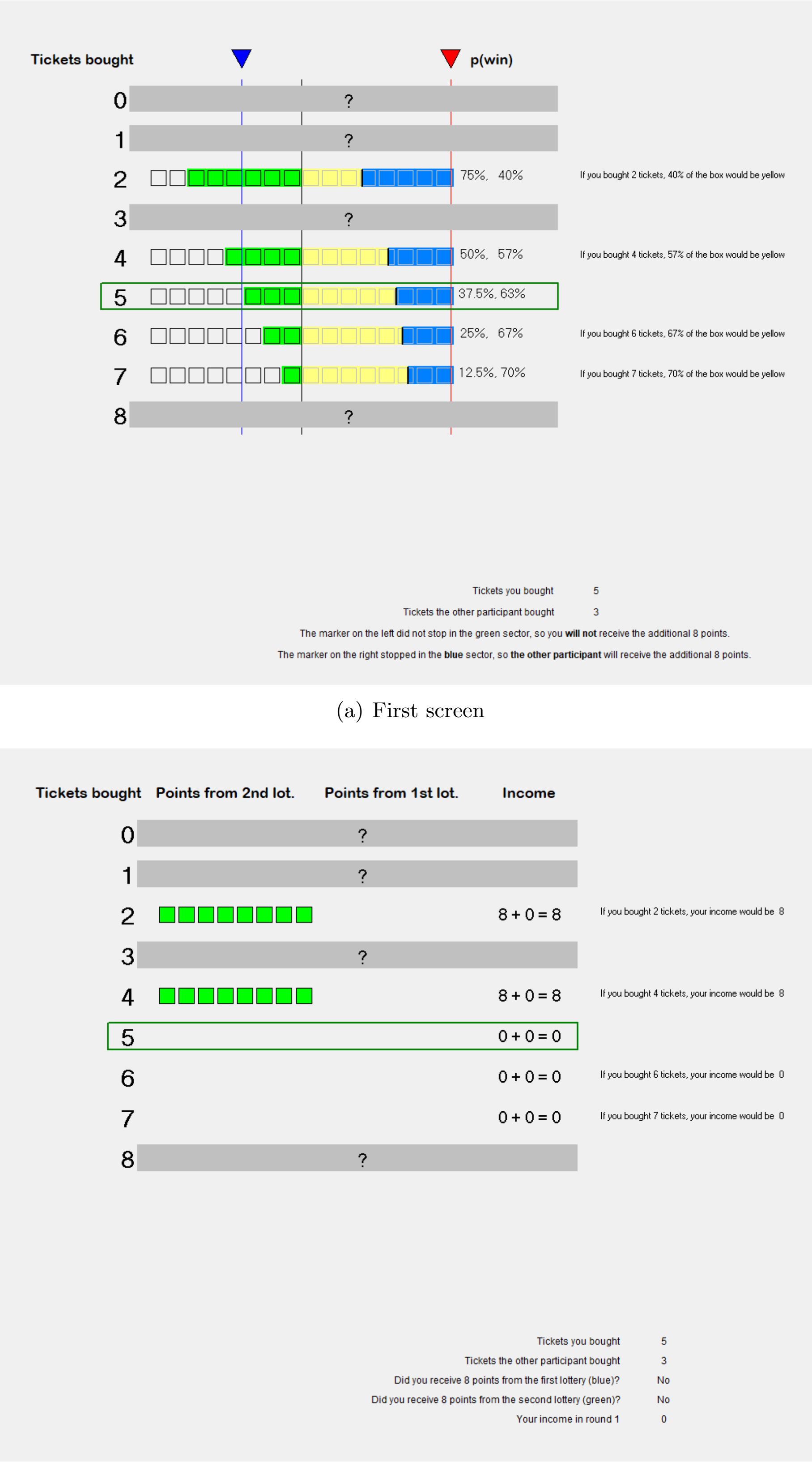
Income received from sharing the additional 8 points (yellow): 5.00

Did you receive the additional 8 points (green) from the lottery? Yes

Your income in round 1 13.00

(b) Second screen

Figure 21: Feedback in RS treatment with FPI.



Tickets bought

p(win)

Tickets bought			
0	?		
1	?		
2		75%, 40%	If you bought 2 tickets, 40% of the box would be yellow
3	?		
4		50%, 57%	If you bought 4 tickets, 57% of the box would be yellow
5		37.5%, 63%	
6		25%, 67%	If you bought 6 tickets, 67% of the box would be yellow
7		12.5%, 70%	If you bought 7 tickets, 70% of the box would be yellow
8	?		

Tickets you bought	5
Tickets the other participant bought	3

The marker on the left did not stop in the green sector, so you **will not** receive the additional 8 points.

The marker on the right stopped in the **blue** sector, so **the other participant** will receive the additional 8 points.

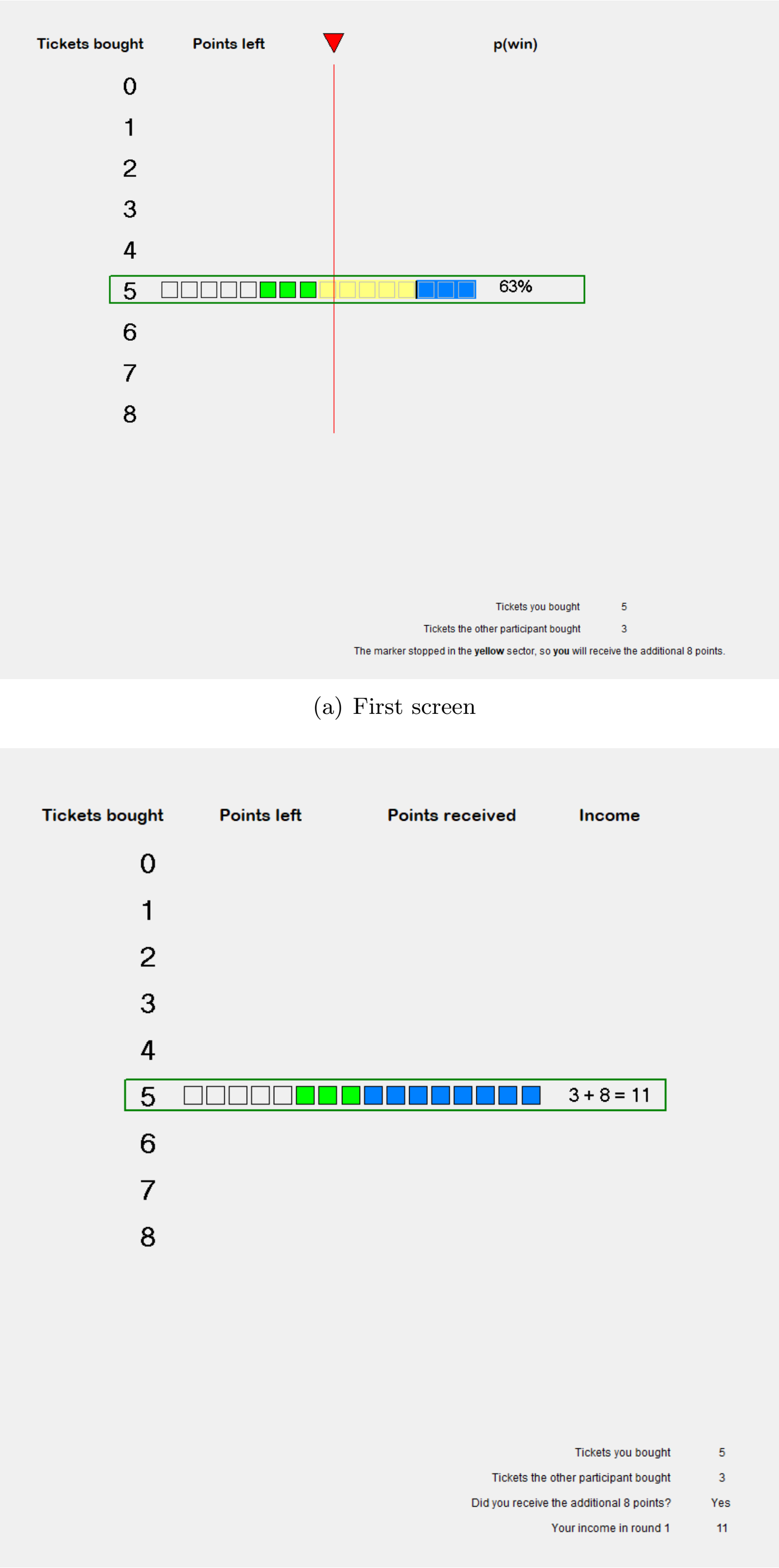
(a) First screen

Tickets bought	Points from 2nd lot.	Points from 1st lot.	Income	
0	?			
1	?			
2	🟩🟩🟩🟩🟩🟩🟩🟩		8 + 0 = 8	If you bought 2 tickets, your income would be 8
3	?			
4	🟩🟩🟩🟩🟩🟩🟩🟩		8 + 0 = 8	If you bought 4 tickets, your income would be 8
5			0 + 0 = 0	
6			0 + 0 = 0	If you bought 6 tickets, your income would be 0
7			0 + 0 = 0	If you bought 7 tickets, your income would be 0
8	?			

Tickets you bought	5
Tickets the other participant bought	3
Did you receive 8 points from the first lottery (blue)?	No
Did you receive 8 points from the second lottery (green)?	No
Your income in round 1	0

(b) Second screen

Figure 22: Feedback in RR treatment with FPI.



Tickets you bought 5

Tickets the other participant bought 3

The marker stopped in the **yellow** sector, so **you** will receive the additional 8 points.

(a) First screen

Tickets you bought	5
Tickets the other participant bought	3
Did you receive the additional 8 points?	Yes
Your income in round 1	11

(b) Second screen

Figure 23: Feedback in SR treatment with no FPI.

I Instructions

Below we reproduce the instructions for SR treatment, with changes in other treatments marked in brackets.

INSTRUCTIONS

Welcome to the experiment. Please read these instructions carefully. They are identical for all the participants with whom you will interact during this experiment. If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions. From now on communication with other participants is not allowed. If you do not conform to these rules you will be excluded from the experiment with no payment. Please also switch off your mobile phone at this moment.

In this experiment you can earn some money. How much you earn depends on your decisions and the decisions of the other participants. During the experiment we will refer to points instead of euros. The total amount of points that you will have earned during the experiment will be converted into euros at the end of the experiment and paid to you in cash confidentially. The conversion rate that will be used to convert your points into your cash payment will be **1 point = 0.45 euros.**

At the end of the experiment you will see how many points you earned in each round. One round from each 10 rounds will be randomly chosen for payment: the first round will be chosen from rounds 1-10, the second round from rounds 11-20 and so on. Your earnings from the selected rounds will be added, converted into euros and paid to you in private at the end of the experiment.

The experiment will contain several parts. Now we will describe you the task that you will be doing in Part 1. At the start of each other part additional instructions will be displayed on your computer screen. Please read those instructions before you continue.

The task in Part 1

In Part 1 there will be 10 rounds. At the beginning of each round the computer will randomly match you with another participant in this room. The participant you are matched with will be changed randomly each round. All participants with whom you will interact will receive the same information and will face exactly the same task.

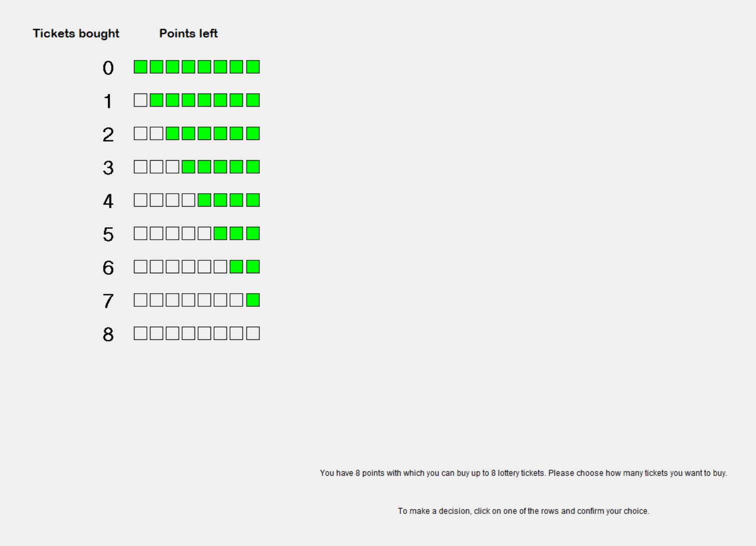
Decision screen

In each round you will receive an endowment of 8 points. You can use these points to buy "lottery tickets" [*SS, RS:* "tokens"]. Each ticket [*SS, RS:* token] you buy costs 1 point, so you can buy up to 8 tickets [*SS, RS: tokens*] each round. Any points that you do not spend on tickets [*SS, RS: tokens*] will be added to your round income [*RS, RR:* will determine the probability to receive points in a second lottery.].

How your decision screen will look like is shown in Figure 1. The 8 points that you receive at the start of each round will be represented by 8 green squares. You will have to choose how many of these points to use to buy lottery tickets [*SS, RS:* tokens]. Every additional ticket [*SS, RS:* token] you buy will take away one green square. To make a decision, click on one of the rows and confirm your choice.

[*SR, RR:* **The lottery**]

[*SR, RR*: After you and the other participant have chosen how many tickets to buy, either you or the other participant will receive additional 8 points. Who receives these additional 8 points will be determined by a lottery. The lottery will be implemented the following way: after you have chosen how many tickets to buy, a box containing 8 squares will be displayed



Tickets bought	Points left
0	🟩🟩🟩🟩🟩🟩🟩🟩
1	□🟩🟩🟩🟩🟩🟩🟩
2	□□🟩🟩🟩🟩🟩🟩
3	□□□🟩🟩🟩🟩🟩
4	□□□□🟩🟩🟩🟩
5	□□□□□🟩🟩🟩
6	□□□□□□🟩🟩
7	□□□□□□□🟩
8	□□□□□□□□

You have 8 points with which you can buy up to 8 lottery tickets. Please choose how many tickets you want to buy.

To make a decision, click on one of the rows and confirm your choice.

Figure 1. Screenshot of the decision screen. [In RS and SS figure showed "Tokens bought"]

on your screen. This box represents the additional 8 points that you may receive. The box will be divided into two sectors, yellow and blue. The yellow sector belongs to you and the blue sector belongs to the other participant. Once you start the lottery, a marker will move around the box and stop at one place at random. The marker is equally likely to stop at any place. If the marker stops in the yellow sector, you will receive the additional 8 points. If the marker stops in the blue sector, the other participant will receive the additional 8 points. The size of the yellow sector represents the probability that you will receive the additional 8 points: if x% of the box is colored yellow, the probability that you will receive the 8 additional points is x%.]

[*SS, RS*: After you and the other participant have chosen how many tokens to buy, you and the other participant will share additional 8 points. These additional 8 points will be displayed as a box, divided into two sectors, yellow and blue. The yellow sector represents your share and the blue sector represents the share of the other participant.]

The sizes of yellow and blue sectors are proportional to the number of lottery tickets [*SS, RS:* tokens] that you and the other participant buy. For example, if you buy the same number of tickets [*SS, RS:* tokens] as the other participant, the yellow sector will be of the same size as the blue sector. If you buy twice as many tickets [*SS, RS:* tokens] as the other participant, the yellow sector will be two times larger than the blue sector. Overall, the probability that you will receive the additional 8 points [*SS, RS:* the share of the additional 8 points that you will receive], represented by the size of the yellow sector, is calculated as follows:

[*SR, RR:*]

$$\begin{array}{l}\text{Probability of receiving} \\ \text{the additional 8 points}\end{array} = \frac{\text{Number of tickets you bought}}{\begin{array}{l}\text{Number of tickets you bought + Number} \\ \text{of tickets the other participant bought}\end{array}} \times 100\%$$

[*SS, RS:*]

$$\begin{array}{l}\text{Share of the additional} \\ \text{8 points}\end{array} = \frac{\text{Number of tokens you bought}}{\begin{array}{l}\text{Number of tokens you bought + Number} \\ \text{of tokens the other participant bought}\end{array}} \times 100\%$$

If nobody buys any tickets [*SS, RS: tokens*], each of you will have a 50% probability to receive the additional 8 points [*SS, RS: will receive 50% of the additional 8 points, that is 4 points*].

[*RS:* **The lottery**] [*RR:* **The second lottery**]

[*RS, RR:* In addition to the share of 8 points, a lottery [*RR:* In addition to the first lottery, a second lottery] will be carried out in which you will have a chance to receive additional 8 points. In this lottery, the probability that you receive the additional 8 points will be determined by the number of points that were not used to buy tokens. On the computer screen these remaining points will be represented by green squares. Each remaining point increases the probability to receive the additional 8 points by 12.5%, as shown in the table:

Points remaining	0	1	2	3	4	5	6	7	8
Probability to receive 8 points	0%	12.5%	25%	37.5%	50%	62.5%	75%	87.5%	100%

The lottery will be implemented as follows: a marker will move and stop at random either on a green square, or on an empty square. If the marker stops on the green square, you will receive additional 8 points. If it stops on an empty square, you will receive 0 points, and the additional 8 points will not be given to anyone. If the lottery determines that you receive additional 8 points, 8 green squares will be added to your round income.] [*RS:* If the lottery determines that you receive additional 8 points, 8 green squares will be added to your round income.]

[*RR:* The two lotteries that are played are independent, so you may receive a total of 16 points, 8 points or 0 points.]

Round income

[*SR:* If the lottery determines that **you** receive the additional 8 points, 8 blue squares from the box will be added to your round income. Your round income will be equal to the number of points that you did not spend buying tickets (green squares) plus the additional 8 points (blue squares). If the **other participant** receives the additional 8 points, the squares from the blue box will disappear and your round income will consist of the points that you did not spend on buying tickets (green squares).]

[*RR:* If the first lottery determines that **you** receive the additional 8 points, 8 blue squares from the box will be added to your round income. Your round income will be equal to the number of points that you did not spend buying tickets (green squares) plus the additional 8 points (blue squares). If the first lottery determines that the **other participant** receives the additional 8 points, the squares from the blue box will disappear. and your round income will consist of the points that you did not spend on buying tickets (green squares). If the second lottery determines that you receive additional 8 points, 8 green squares will be added to your round income.]

[*SR:*

- If you receive the additional 8 points, your round income is:

  Round Income = 8 —- Number of purchased tickets + 8

- If you do not receive the additional 8 points, your round income is:

  Round Income = 8 —- Number of purchased tickets]

[*RS:*

- If you receive the additional 8 points from a lottery, your round income is:

  Round Income = (Share of the additional 8 points) * 8 + 8

- If you do not receive the additional 8 points from a lottery, your round income is:

  Round Income = (Share of the additional 8 points) * 8]

[*RR:*

- If you receive the additional 8 points in the first lottery and in the second lottery, your round income will be 16 points.
- If you receive the additional 8 points only in one of the lotteries, your round income will be 8 points.
- If you do not receive the additional 8 points from either lottery, your round income will be 0 points.]

[*SS:* Your round income will be equal to the number of points that you did not spend buying tokens (green squares), plus the share of the additional 8 points (yellow sector):

Round income = 8 —- Number of purchased tokens + (Share of the additional 8 points) * 8]

At the end of a round you will see the number of tickets [*SS, RS: tokens*] the other participant bought, your probability to receive [*SS, RS: your share of*] the additional 8 points, [*SR, RS:* the outcome of the lottery] [*RR:* the outcome of the lotteries] and your round income.

End of the experiment

At the end of the experiment you will be informed about your income in the rounds that were randomly selected for payment. Income from these rounds will be added, converted into euros and paid in private once you complete a short questionnaire. Please stay seated until we ask you to come to receive the earnings.

If you have any further questions, please raise your hand now. If you have read the instructions and have no further questions, please click “Start the Experiment” on your computer screen.

Before starting Part 1 of the experiment we will ask you to complete two trial rounds. These two rounds will be the same as the task in Part 1, but income in these two rounds will not affect your final earnings.

Notes:

- The participant you are matched with will be randomly changed each round.
- One round from each 10 rounds will be randomly selected for payment.
- Part 1 will have 10 rounds. The number of rounds in other parts and additional instructions will be displayed on your computer screen.
- You can buy tickets [*SS, RS:* tokens] only using your endowment, which is equal to 8 points in each round.

I.1 Additional on-screen instructions

At the start of block 2, participants saw additional instructions on the computer screen, informing about the availability of foregone payoff information:

This is the start of **Part 2**.

In Part 2 there will be **20** rounds.

Two rounds from these 20 will be randomly chosen for payment at the end of the experiment.

The task you do will be the same as the task you did in Part 1. The only difference is that at the end of each round you will have an option to reveal information about what would have happened if you had bought a different number of [*SR, RR*: lottery tickets; *SS, RS*: tokens]. To uncover this information, click on any grey box marked with "?". Then the box will disappear and you will see what would have been the sizes of yellow and blue sectors [*SS:* and your round income] if you had bought a different number of [*SR, RR*: tickets; *SS, RS*: tokens]. [*SR, RS, RR:* You will also see what would have been the outcome of the lottery [*RR*: two lotteries] and whether or not you would have received the additional 8 points.] This additional information will have no effect on your earnings.

If you want to consult these instructions during the experiment, please raise your hand and the experimenter will bring you a paper copy.