University of Innsbruck

Hot hand and gambler's fallacy in teams: Evidence from investment experiments

Thomas Stöckl, Jürgen Huber,
Michael Kirchler, Florian Lindner

Working Papers in Economics and Statistics

2013-04

Research Platform eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact Address:
University of Innsbruck
Department of Public Finance
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Hot Hand and Gambler's Fallacy in Teams: Evidence from Investment Experiments*

Thomas Stöckl† Jürgen Huber‡
Michael Kirchler§ Florian Lindner¶

August 26, 2014

Abstract

In laboratory experiments we explore the effects of communication and group decision making on investment behavior and on subjects' proneness to behavioral biases. Most importantly, we show that communication and group decision making do not impact subjects' overall proneness to the hot hand fallacy and to the gambler's fallacy. However, groups decide differently than individuals, as they rely significantly less on useless outside advice from "experts" and choose the risk-free option less frequently. Furthermore we document gender differences in investment behavior: groups of two female subjects choose the risk-free investment more often and are marginally more prone to the hot hand fallacy than groups of two male subjects.

JEL classification: C91, C92, D81, G10
Keywords: hot hand fallacy; gambler's fallacy; experimental finance; team decision making.

---

*Financial Support from OeNB grant # 14953 is gratefully acknowledged.
†Corresponding author. Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria e-mail: thomas.stoeckl@uibk.ac.at.
‡Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria.
§Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria and Centre for Finance, Department of Economics, University of Gothenburg, Vasagatan 1, 41124 Gothenburg, Sweden.
¶Innsbruck University, Department of Public Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria.

1 Introduction

The hot hand fallacy and the gambler's fallacy are two important behavioral biases in financial markets. People who are affected by these biases misinterpret random sequences. Specifically, when prone to the hot hand fallacy people misidentify a non-autocorrelated sequence as positively-autocorrelated, generating beliefs that a run of a certain realization will continue in the future. In financial markets this bias is e.g. observable when investors delegate decisions to experts like professional fund managers. Specifically, people mostly buy funds which were successful in the past, believing in the managers' ability to prolong the performance record (see e.g. Sirri and Tufano, 1998; Barber et al., 2005). Rabin (2002) calls this phenomenon overinference.

With the gambler's fallacy people expect that, even in a short sequence of events, possible realizations should be represented according to the overall probabilities (Tversky and Kahneman, 1971). Expressed more formally: a non-autocorrelated random sequence is believed to exhibit negative autocorrelation. The disposition effect can be seen as an exhibition of the gambler's fallacy, as investors (private and institutional alike) sell winners too soon and hold losers too long (Odean, 1998; Weber and Camerer, 1998; Shapira and Venezia, 2001; Rabin, 2002; Chen et al., 2007). Kroll et al. (1988) document sequential dependencies, predominantly the gambler's fallacy in a portfolio selection task.

Biased decisions can lead to unfavorable or negative consequences for the decision maker. For instance, Goetzmann and Kumar (2008) document that U.S. investors who exhibit trend-related behavior – either trend chasing (hot hand) or contrarian (gambler's fallacy) – hold less diversified portfolios, implying negative risk and performance consequences. Investors' belief in hot hands of mutual fund managers (Brown et al., 1996; Chevalier and Ellsion, 1997; Sirri and Tufano, 1998) generates fund inflows negatively related to the past rank of a mutual fund. However, given the lack of persistence in fund performance (see e.g. Carhart, 1997; Malkiel, 2003, 2005) this behavior leads to biased decisions. In a different context, Dohmen et al. (2009) relate the hot hand fallacy and the gambler's fallacy to an increased probability of long-term unemployment and to a higher probability to overdrawn bank accounts, respectively. Galbo-Jørgensen et al. (2013) use data on lotto gambling and find evidence for both biases. They show that players tend to bet less on numbers that were drawn in the last week (gambler's fallacy) and bet more on a number if it was frequently drawn in the recent past (hot hand fallacy). Given the negative consequences of biases for those prone to them, it is important to focus on potential strategies to neutralize them.

By using investment experiments Huber et al. (2010) investigate both biases in a unified framework. Participants in their experiment are confronted with a series of independent coin tosses showing head and tail with probability 0.5 each. They can choose to (a) predict the realization of the next coin toss themselves, (b) delegate the decision to computerized random agents, called experts, or (c) take a risk-free payment. As reward subjects receive 100 Taler (the experimental currency) for a correct decision while 50 Taler are deducted for an incorrect one. Delegating the investment decision to an expert offers the same payoffs, but a fee is deducted. The risk-free option offers a reward of 10 Taler with certainty. Hence, payoffs are calibrated such that predicting for oneself is preferred to delegating the decision to an expert and the latter is preferred to the risk free alternative, for a participant who is risk neutral (with the implicit assumptions that (i) they believe the coin toss is i.i.d., (ii) they believe the coin has a 50% chance of heads, and (iii) they understand how to optimize in this environment).

Huber et al. (2010) observe both, the hot hand and the gambler's fallacy, in subjects' decisions. Specifically, experts are selected more frequently, the more successful they had been in the past. This implies that subjects expect hot hands in the computerized agents' decisions. In addition, among subjects picking head or tail themselves the authors observe the gambler's fallacy as head (tail) is chosen less frequently after streaks of heads (tails).[1] By using a similar framework but labelling experts differently, Powdthavee and Riyanto (2014) report strong hot hand fallacies to outside advice for the outcome of randomized coin tosses. In their paper "experts" were modelled as envelopes with predetermined advice for each period of the investment game.

Here we use the setup of Huber et al. (2010) to study the effects of team decision making on investment decisions and behavioral biases. Many, probably most, decisions of huge economic importance are made by groups rather than individuals – for instance, the "Federal Open Market Committee" of the FED consists of seven members and the "Governing Council of the European Central Bank" consists of 23 members that jointly decide on monetary policy. In financial markets, teams of fund managers decide on the investment strategy of a fund and which stocks to pick.[2] Ample evidence in the literature supports the positive impact of group decision making on decision quality. Irrespective of decisions being made in strategic or non-strategic situations, groups usually

[1] Ackert et al. (2012) report that hiding information of past realizations prevents subjects in their experiment from exhibiting the gambler's fallacy in portfolio decision experiments. This approach, however, seems practically impossible, given the large amount of available financial data and the attention this data generates.

[2] Bär et al. (2011) document that teams of fund managers implement less extreme investment styles and less industry concentrated portfolios. In an experiment Rockenbach et al. (2007) find that team decisions are better in line with Portfolio Selection Theory than individual decisions, leading to a better risk-return ratio. Keck et al. (2014) demonstrate that groups are more likely than individuals to make ambiguity-neutral decisions. They attribute this to effective communication in groups.

perform equally well or better than individuals.[3] Though group decision procedures are widely implemented, we know surprisingly little about how they affect potentially present behavioral biases in financial markets.[4]

We focus on two research questions (RQ). In RQ 1 we analyze differences in decision making between individuals and groups on the aggregate level and over time. In a second step, we split our sample to investigate potential effects originating from the gender composition of groups. The second part of RQ 1 is motivated by ample previous literature highlighting differences in decision making by gender, which we also expect to play a role in our setting.[5]

> **RQ 1**: Do groups decide differently compared to individuals in selecting their investment or in relying on outside advice? Does gender composition of groups play a role?

In RQ 2 we investigate whether group decision making leads to financial decisions less influenced by behavioral biases. The second part of RQ 2 again focuses on whether gender effects behavioral biases. Dohmen et al. (2009) and Suetens and Tyran (2012) provide some (inconclusive) evidence on this issue.

> **RQ 2**: Are groups differently prone to behavioral biases such as the gambler's fallacy and the hot hand fallacy compared to individuals? Does gender composition of groups play a role?

The presence of behavioral biases in the investment decision experiment of Huber et al. (2010) allows us to test the robustness of their results for group decision making. Using Treatment INDIV as reported in Huber et al. (2010) as our benchmark, we conduct further treatments with different levels of group decision making. In treatments COMM and GROUP subjects are assigned to groups of two and a chat is installed. While communication is possible in both treatments they differ in the way decision making takes place. In Treatment COMM subjects can communicate, but decide individually. In Treatment GROUP subjects have to agree on a decision as a group.

We find that (i) communication and group decision making does not impact subjects' overall proneness to behavioral biases like gambler's fallacy and hot hand fallacy. (ii) However, groups in Treatment GROUP rely less on useless expert advice compared to the other treatments. (iii) Group decision making

[3]Evidence in strategic games is provided in Feri et al. (2010), Sheremeta and Zhang (2010), Cheung and Coleman (2011), Casari et al. (2012) and Sutter et al. (2013). Evidence in non-strategic games is provided in Bone et al. (1999), Blinder and Morgan (2005), Charness et al. (2007), Rockenbach et al. (2007), Sutter (2007) and Fahr and Irlenbusch (2011). See Charness and Sutter (2012) and Kugler et al. (2012) for comprehensive reviews.

[4]Charness et al. (2010) demonstrate that the conjunction fallacy is diminished substantially when groups of two or three communicate before making a decision. In an investment game Sutter (2009) finds no difference between individual and team decisions.

[5]See Croson and Gneezy (2009) for a review of gender differences in economic experiments.

in Treatment GROUP leads to fewer choices of the risk-free alternative and to more own guesses on the realization of the coin toss compared to the other treatments. (iv) Finally, we observe that gender composition of groups plays a crucial role in investment behavior: groups of two female subjects choose the risk-free investment significantly more often and delegate investment decisions less often to experts than groups of two male subjects. In addition, we are the first to document that women (INDIV) and female-only groups (COMM and GROUP) show a marginally higher proneness to the hot hand fallacy.

This paper is structured as follows: In Section 2 the design of the decision problem and the treatments are outlined. Section 3 describes the conceptual framework, Section 4 presents the results and Section 5 summarizes and discusses the results.

2 The experiment

2.1 Design of the decision problem

At the beginning of the experiment subjects receive an initial endowment of 500 Taler (the experimental currency). In each of 40 periods subjects are asked to choose between a risky and a risk-free investment which differ in payouts (see Figure 1).

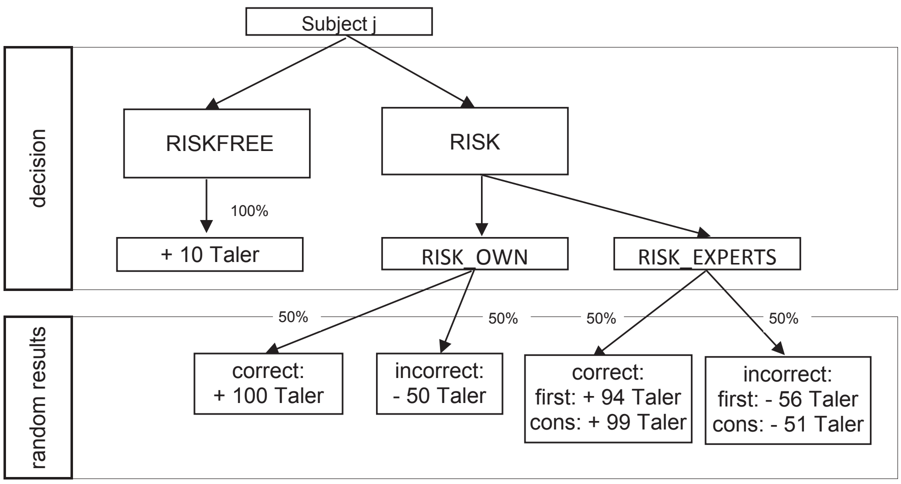


Figure 1: Design of the decision problem and payouts for one period.

When selecting the risk-free alternative (RISKFREE) subjects earn 10 Taler with certainty. The risky investment is simulated by a coin toss showing head and tail with equal probabilities. The subjects' task when going for this alternative is to choose one side of the coin. This can be done in two distinct ways.

First, subjects can make own guesses on the realization of the coin (RISK_{OWN}) or, second, they can delegate the decision to one of five computerized agents – labelled "experts" ($\text{RISK}_{\text{EXPERT}}$), who then randomly pick one side of the coin for the subject.[6]

If subjects make own guesses on the realization of the coin toss, they earn 100 Taler if their guess coincides with the random coin realization, otherwise they lose 50 Taler, for an expected profit of 25 Taler. When subjects delegate decisions to the experts they have to pay two types of fees. First, an issue surcharge of 5 Taler is deducted if subjects select an expert that they did not choose in the previous period. Staying with the same expert in the following periods does not trigger the fee again. Second, a management fee of 1 Taler is collected each period a subject selects one of the experts.[7] If the expert's decision and the coin realization are identical, 100 Taler minus charges are added to the subjects' account. In the opposite case, 50 Taler plus the charges are subtracted from his account (see Figure 1).

Taking a look at the payouts of the risky investment note that RISK_{OWN} dominates $\text{RISK}_{\text{EXPERT}}$. RISK_{OWN} exhibits a higher expected payout value and offers superior payouts for each state of nature (win, lose) as no fees apply. While the choice between the riskfree option and the risky alternatives is subject to risk aversion, the choice of $\text{RISK}_{\text{EXPERT}}$ clearly constitutes an "inferior" choice compared to RISK_{OWN}.

2.2 Treatments

In Treatment INDIV each subject chooses between RISK_{OWN}, $\text{RISK}_{\text{EXPERT}}$, and RISKFREE individually. No communication between subjects is allowed and actions by one subject do not influence actions or outcomes of other subjects.

In Treatment COMM subjects are randomly paired at the beginning of the experiment. Pairs are kept unchanged for the entire experiment. The two subjects in a pair can chat for up to 90 seconds each period before making a decision.[8] The chat area is placed in the lower third of the main screen such

[6] One could argue that the computerized coin toss as well as the experts' decisions generate (in fact false) beliefs of experimenter manipulations. Also, the term "experts" might suggest superior decision skills and attract subjects' attention. Both concerns generate ambiguity with respect to subjects' prior expectations about randomness and experts' skills. However, in a setting similar to ours Powdthavee and Riyanto (2014) tackle both concerns and explore potential consequences. In their experiment they rule out experimenter manipulation by tossing the coin in public and distributing experts' decisions at the beginning of the experiment. Furthermore, they refrain from labelling advice as "experts." Most importantly, these changes do not affect subjects' decision making or proneness to biases.

[7] This structure is similar to what many investment funds charge, i.e., an issue (entrance) surcharge and then an annual management fee.

[8] The chat time was reduced to 60 seconds after period 15 and it can be ended any time

that subjects can access their past decisions, their performance, and the experts' performance anytime during the experiment (see Appendix B of the online supplement for instructions and screenshots.) While subjects can exchange information and expertise via the chat, their decisions are still individual decisions, i.e., the decision of the chat-partner has no influence on the subject's payout.[9]

In Treatment GROUP a chat is set up exactly as in Treatment COMM, but now the chat-partners are incentivized to reach a joint decision. While each subject still has to enter his decision individually on the screen, they can only earn a positive payoff if they select the same investment. If the chat-partners' decisions are identical, payouts are calculated as previously specified. However, when the decisions of the two chat-partners are not identical they are redirected to the chat for another 45 seconds. If the newly entered choices are still inconsistent, subjects are penalized by deducting 50 Taler from each subjects' account irrespective of their choices.[10]

2.3 Implementation of the experiment

During the experiment each subject has access to several sources of information on the trading screen (see screenshots in Appendix B of the online supplement). His current wealth, the number of periods played, previous decisions, the current and past realizations of the coin, his success/failure, and the changes of his holdings in each period are displayed. Furthermore, subjects are informed about the past history of the experts. In the starting period subjects see a randomly generated series of five previous (imaginary, i.e., not played) periods (periods t-4 to t=0) of the experts' history.[11] In the lower part of the screen a performance measure for experts is presented, which displays the percentage of correct decisions within the previous five periods.

The realizations of the coin tosses are drawn randomly in advance and we use the same realizations for each session to ensure comparability across sessions. For a detailed list of the coin realization and information about the experts' performance in each period see Table A1 of Appendix A in the online supplement.

We conducted 18 sessions (6 per treatment) with a total of 360 subjects

before the official stop by clicking an "End Chat"-Button.

[9] Of 2,400 decision pairs in Treatment COMM 1,113 (46.4%) were different between the two subjects of a group, while 1,287 (53.6%) were identical.

[10] We chose this design to make clear to subjects that they need to reach a joint decision. Out of 2,400 decisions in Treatment GROUP subjects did not reach a joint decision in only five cases (0.2%), four of which happened in periods 1 and 2.

[11] This sort of information is easily accessible on real financial markets as it is an important marketing tool of mutual funds. Kroll et al. (1988) document a high demand for past return realizations in their experiment, though the knowledge of the underlying process reveals the uselessness of this sort of information.

(120 per treatment).[12] In total we observed 4,800 decisions per treatment yielding a total of 14,400 decisions to analyze. The experiments were conducted with z-Tree (Fischbacher, 2007) and took place at the University of Innsbruck. Treatment INDIV, taken from Huber et al. (2010), was run in March 2006 while treatments COMM and GROUP were run in June 2009.[13] Subjects were recruited using ORSEE by Greiner (2004).

At the end of the experiment subjects' accumulated Taler holdings were exchanged into Euros at a known fixed rate of 100:1 and paid out privately in cash. The average payout was EUR 14.

3 Conceptual framework

We develop a conceptual framework to model the hot hand and gambler's fallacy following the approach of Powdthavee and Riyanto (2014, pages 13–16). This model is a simplified version of the model presented in Rabin and Vayanos (2010) applicable to the decision problem in our experiment.

In line with Rabin and Vayanos (2010), we initially assume that each participant in the experiment observes two sequences of informative (public) signals whose probability distributions depend on some underlying states before deciding between RISK_{OWN} and $\text{RISK}_{\text{EXPERT}}$ in period t. The first signal, s_t, represents the realization of a (fair) coin toss in periods $t = 1, 2, ..., 40$, with 1 signalling "head" and 0 signaling "tail":

$$s_t = \mu + \epsilon_t. \tag{1}$$

The second signal a_t provides the expert's prediction in period t with a value of 1 if the prediction of an expert matches the outcome of the coin realization and 0 otherwise:[14]

[12] Tables A4 and A5 in Appendix A of the online supplement provide details on demographic characteristics (age, gender, semester of studying) and subjects' answers to questions about overconfidence, stock market experience, and mood across treatments and for each single session. We find no significant differences between treatments and no systematic differences between sessions.

[13] The attentive reader recognizes that INDIV was run before the financial crisis, while COMM and GROUP were run after the climax of the financial crisis in 2008 when Lehman Brothers went bankrupt on September 15. The considerable time lag between treatments might raise concerns about the interpretability of treatment effects because the financial crisis may well have affected how people think about "experts". However, we are convinced that our student subject pool was unlikely to have different priors in 2006 than in 2009. This argument is supported by Figure 2 revealing that the share of subjects choosing an expert at the beginning of the experiment is highest in GROUP (57%), closely followed by COMM (53%), but markedly lower at 36% in INDIV. If the crisis had shaken trust in experts in our subject pool, we should observe just the opposite. Note that this concern does not restrict interpretation of treatment effects between COMM and GROUP.

[14] In contrast to the experiment of Powdthavee and Riyanto (2014), in our setting both sets of signals are identical and public for all subjects.

$$a_t = \phi + u_t, \tag{2}$$

where μ represents the long-run mean of the i.i.d. signals, which is obviously 0.5 for a fair coin and *a priori* fixed in our experiment. Due to the fact that an expert faces the same fair coin, the expert's long-run prediction (ϕ) theoretically equals 0.5.[15] ϵ_t and u_t are i.i.d. normal shocks with zero means and variances strictly greater than zero. One interpretation of the shock u_t (ϵ_t) is the luck of an expert (a subject) correctly predicting s_t in period t. Note, that it is assumed that s_t and a_t are to be determined independently, indicating that coin realizations and experts' predictions in a certain period do not influence each other.

In this context behavior consistent with the gambler's fallacy arises when subjects have a mistaken belief about the sequence of the normal shocks (ϵ_t, u_t) not being i.i.d. but exhibiting systematic reversal (see Rabin and Vayanos, 2010).[16] This false perception implies that (i) subjects will develop an erroneous belief that the coin realization in period t is more likely to be tail $(1 - s)$ after a streak of heads (s) up to $t-1$, and (ii) subjects will develop an erroneous belief that an experts' prediction in period t is more likely to be incorrect $(1-a)$ following a streak of correct predictions (a) up to $t - 1$.

By allowing for subjects' perception about the nature of ϕ to be influenced by a streak of correct (a) or incorrect predictions $(1 - a)$ up to $t - 1$, it becomes possible to model behavior consistent with the hot hand, overruling the gambler's fallacy. More precisely, subjects' perception about an experts long-run ability to predict the coin realization can change from being fixed at 0.5 to one that is developed according to the auto-regressive process:

$$\phi_t = 0.5 + \rho(\phi_{t-1} - 0.5) + \eta_t, \tag{3}$$

where $0 \leq \rho_i < 1$ stands for the reversion rate to the long-run average of 0.5, η_t is an i.i.d. normal shock with zero mean, variance greater than zero, and independent of u_t. For $\rho > 0$ a belief in a serially correlated variation in ϕ can evolve (i.e., a belief in hot hand).[17,18]

[15] The *a priori* randomly drawn predictions of the experts 1 to 5 lead to an *ex post* success rate of 0.45, 0.525, 0.525, 0.4, and 0.375, respectively.

[16] See also Rabin (2002) who uses a different approach to model false beliefs in the law of small numbers and Asparouhova et al. (2009), who show, using structural estimation, that this models generates the best fit in an experiment testing beliefs in regime shifting and the law of small numbers.

[17] Note that the hot hand and the gambler's fallacy are not symmetric concepts.

[18] In principal, this could also lead to a cold hand, in case of a sequence of incorrect predictions.

4 Results

4.1 Investment decision quality

To tackle RQ 1 on the effects of group decision making on investment decisions, we first compute for each subject/group the ratio of decisions for predicting the coin toss on her/their own (RISK_{OWN}), the ratio of delegated decisions to experts ($\text{RISK}_{\text{EXPERT}}$), and the ratio of the risk-free alternative (RISKFREE). Thus, the number of observations equals 120 in INDIV and COMM and 60 in GROUP. Treatment averages (column 3) and averages by gender composition (columns 4-6) are outlined in Table 1. We apply Mann-Whitney U tests to determine the statistical significance of treatment and gender effects.

Table 1: Investment decisions across treatments and gender. RISK_{OWN} stands for the ratio of subjects/groups predicting the realization of the coin flip on their own. $\text{RISK}_{\text{EXPERT}}$ measures the ratio of delegated decisions to experts among all decisions. RISKFREE indicates the ratio of choices for the risk-free alternative. M (F) denotes male (female) individuals, MM denotes male only groups, MIX are mixed groups, and FF are female only groups.

Treatment	Decisions	All	M/MM	MIX	F/FF
INDIV	RISK_{OWN}	68.8%	67.6%		70.7%
	$\text{RISK}_{\text{EXPERT}}$	23.8%	27.1%		18.4%
	RISKFREE	7.5%	5.3%		10.9%
COMM	RISK_{OWN}	71.8%	67.9%	75.4%	66.9%
	$\text{RISK}_{\text{EXPERT}}$	23.6%	29.2%	20.5%	22.6%
	RISKFREE	4.7%	2.9%	4.1%	10.4%
GROUP	RISK_{OWN}	79.4%	80.6%	78.7%	79.5%
	$\text{RISK}_{\text{EXPERT}}$	17.2%	18.6%	18.4%	13.6%
	RISKFREE	3.4%	0.8%	2.9%	6.9%

In each treatment the majority of decisions is observed in category RISK_{OWN}. However, compared to the benchmark treatment (INDIV) we notice that in treatments COMM and GROUP choices for RISK_{OWN} are 3.0 and 10.6 percentage points higher, respectively. While the impact of communication is small and insignificant, the marked difference between INDIV and GROUP reveals a significant shift in the decision behavior between individuals and groups with groups being closer to an expected payout maximizing strategy (Mann-Whitney-U test, p-value=0.0456, N=180). Almost the mirror image emerges for $\text{RISK}_{\text{EXPERT}}$: decisions delegated to experts are on average highest when subjects decide individually. Communication among groups does not significantly impact decision behavior. However, when deciding in groups, experts are chosen less frequently with only 17.2% of decisions delegated to them, compared to 23.8% and 23.6% in INDIV and COMM, respectively. Applying Mann-Whitney U tests, however, these differences turn out insignificant. Choices for RISKFREE are highest in

INDIV where 7.5% of decisions are observed in this category. In treatments COMM and GROUP we see a reduction as only 4.7% and 3.4% of decisions are made for RISKFREE. Compared to the benchmark this constitutes a reduction of 37% and 55%, respectively. GROUP exhibits a significantly lower share of RISKFREE decisions as INDIV (Mann-Whitney U test, p-value=0.0050, N=180) and COMM (Mann-Whitney U test, p-value=0.0275, N=180).[19]

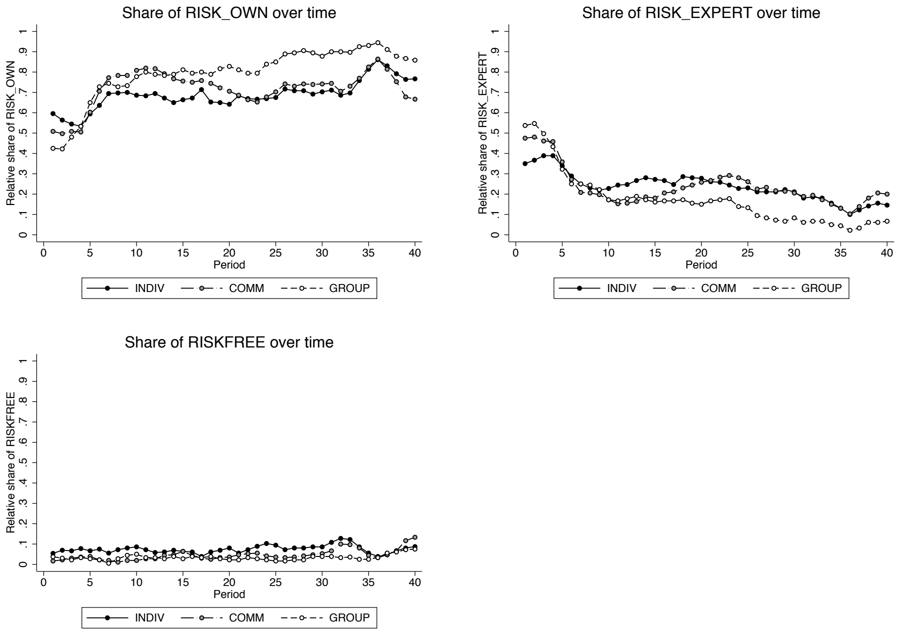


Figure 2: Decisions for RISK_{OWN} (upper left panel), $\text{RISK}_{\text{EXPERT}}$ (upper right panel), and RISKFREE (lower left panel) as percentage shares of all decisions. Each dot represents a moving average over three periods (t-1, t, and t+1).

Next, we analyze subjects' decision behavior in more detail by looking at its development over time. Figure 2 shows 3-period-moving-averages (i.e., an average over periods t-1, t, and t+1) of ratios of subjects/groups choosing RISK_{OWN} (upper left panel), $\text{RISK}_{\text{EXPERT}}$ (upper right panel), and RISKFREE (lower left panel). We find that the share of subjects/groups choosing $\text{RISK}_{\text{EXPERT}}$ is highest in the first couple of periods and markedly lower in the following ones in all treatments. More specifically, in the first period the share of subjects choosing an expert is highest at almost 57% in GROUP and 53% in COMM, but markedly lower at 36% in INDIV. The high initial shares are most likely triggered by subjects' uncertainty about the experts' skills, which are difficult to assess at the beginning of the experiment. Following this initial phase the

[19] Masclet et al. (2009) find that groups are more likely than individuals to choose safe lotteries; however, differing from their study, in our setting the expected value from the risky choice (25 Taler on average) is much higher than the risk-free payout of 10 Taler.

difference between treatments completely vanishes and shares fluctuate between 20 and 24%. In treatments INDIV and COMM the share of subjects choosing $\text{RISK}_{\text{EXPERT}}$ now stabilizes at roughly 21%. Thus, learning seems to have come to an end after periods 5-10 in these treatments indicating that a substantial number of subjects still believe in the experts' skills even after gaining sufficient experience about the experts' performance. A different picture emerges for treatment GROUP in which the learning process continues until the end of the experiment when only 3% choose an expert in the final period. Most interestingly, beliefs in the experts' skills do not completely die out. Evidence on the issue is found be looking at the very last periods. In all treatments the share of subjects/groups choosing $\text{RISK}_{\text{EXPERT}}$ slightly increases at the end of the experiment. This pattern would not have occurred if subjects would consider expert advice useless.

The loss in $\text{RISK}_{\text{EXPERT}}$ in the starting phase of the experiment is compensated by an increase in RISK_{OWN}. After that corrective behavior occurred the share of RISK_{OWN} remains constant in INDIV and COMM but slightly increases in GROUP, mirroring the results observed in $\text{RISK}_{\text{EXPERT}}$. RISK_{OWN} exhibits a decrease in the last periods of the experiment due to an increase in $\text{RISK}_{\text{EXPERT}}$ and RISKFREE. The latter behavior might be explained by subjects trying to shield their earnings from previous periods from potential losses in the final periods.

We now turn to the second part of RQ 1 and further split results by gender in columns 4-6 of Table 1. M/MM denotes male individuals or groups composed of two men; F/FF respectively stands for female individuals or groups and MIX for groups composed of one male and one female participant. The numbers shown in Table 1 reveal no distinct gender effect in the ratio of decisions for RISK_{OWN}. However, groups involving female participation seem to judge the experts more sceptically than groups involving only male participants. We find that MIX groups choose $\text{RISK}_{\text{EXPERT}}$ less frequently as indicated by positive percentage point differences in six out of seven cases. While these results indicate a clear tendency, they borderline conventional significance levels and should thus be interpreted carefully. The risk-free alternative (RISKFREE) is consistently chosen more frequently when female subjects are involved in the decision process. The ratio of RISKFREE is higher in all subgroups and significant in three out of seven cases (Mann-Whitney U tests: M vs. F, p-value 0.013, N=120; MM vs. FF, p-value 0.089, N=32; MIX vs. FF, p-value 0.050, N=44). Thus, our data supports the widespread evidence that female subjects and female-only groups choose less risky options than their male counterparts (see Croson and Gneezy, 2009, and citations therein for a review of evidence).

To strengthen the presented results on decision behavior, time trends, and gender effects we run probit regression (see Table 2) on individual period decision data. Therefore, we regress the individual subjects' binary choices for RISK_{OWN} (1 if RISK_{OWN}, 0 otherwise), $\text{RISK}_{\text{EXPERT}}$ (1 if $\text{RISK}_{\text{EXPERT}}$, 0 otherwise), and RISKFREE (1 if RISKFREE, 0 otherwise) on a constant (α), two treatment dummies for COMM and GROUP, a time trend variable running from 1 to 40 (*Period*), and the variable *Group Comp.* that discriminates individuals/groups according to gender (0=M/MM, 1=MIX, 2=F/FF). This set of regressors constitutes Model 1. To identify differences in learning between treatments we set up a second regression model (Model 2) in which we add two terms interacting the treatment dummies for COMM and GROUP with *Period*. The

Table 2: Probit regressions on subjects' binary choices for RISK_{OWN} (1 if RISK_{OWN}, 0 otherwise), $\text{RISK}_{\text{EXPERT}}$ (1 if $\text{RISK}_{\text{EXPERT}}$, 0 otherwise), and RISKFREE (1 if RISKFREE, 0 otherwise) based on individual period data with standard errors (in parentheses), clustered at the individual level for INDIV and group level for COMM/GROUP.

	Model 1			**Model 2**		
	RISK_{OWN}	$\text{RISK}_{\text{EXPERT}}$	RISKFREE	RISK_{OWN}	$\text{RISK}_{\text{EXPERT}}$	RISKFREE
α	0.112	-0.151	-1.870***	0.199**	-0.272***	-1.756***
	(0.099)	(0.103)	(0.144)	(0.099)	(0.103)	(0.140)
COMM	0.084	0.008	-0.244*	0.170	-0.022	-0.569***
	(0.112)	(0.121)	(0.125)	(0.119)	(0.126)	(0.167)
GROUP	0.341***	-0.221*	-0.448**	-0.038	0.217	-0.517**
	(0.122)	(0.126)	(0.176)	(0.149)	(0.145)	(0.213)
Period	0.018***	-0.024***	0.009***	0.014***	-0.017***	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Group Comp.	0.025	-0.139**	0.264***	0.024	-0.139**	0.264***
	(0.060)	(0.064)	(0.073)	(0.060)	(0.064)	(0.073)
COMM*Period				-0.004	0.001	0.014***
				(0.004)	(0.005)	(0.005)
GROUP*Period				0.020***	-0.025***	0.003
				(0.006)	(0.007)	(0.007)
N	14.400	14.400	14.400	14.400	14.400	14.400
Clusters	240	240	240	240	240	240
R-squared	0.0293	0.0455	0.0460	0.0357	0.0532	0.0487
Prob > $\chi 2$	0.000	0.000	0.000	0.000	0.000	0.000

Notes: COMM and GROUP are treatment dummies; *Period* is a period indicator and runs from 1 to 40; *Group Comp.* distinguishes group composition 0=M/MM, 1=MIX, 2=F/FF. ***, ** and * represent significance at the 1, 5 and 10 percent levels respectively.

regression results confirm our previously reported results. While communication has limited impact on individual decision behavior, group decision making significantly increases probabilities for RISK_{OWN} to be chosen and lowers probabilities for the dominated option $\text{RISK}_{\text{EXPERT}}$. Also RISKFREE is chosen with lower probability. The results on time trends for RISK_{OWN} and $\text{RISK}_{\text{EXPERT}}$ support the graphical findings presented above. We observe a significantly positive (negative) coefficient of *Period* in RISK_{OWN} ($\text{RISK}_{\text{EXPERT}}$). In contrast to the coefficient COMM*Period of Model 2 the coefficient of GROUP*Period turns out significant indicating continuing learning in that treatment. Results for RISKFREE support previous findings on fewer selection in COMM and GROUP but reveal a small but significant time trend (see coefficient of *Period*). The latter effect vanishes in Model 2. We additionally run regressions testing for gender differences in treatments by interacting the treatment dummies for COMM and GROUP with *GroupComp*. We do not report these results here as we do not find any significant influence of these variables while the other coefficient values remain unchanged.[20]

Finally, we study how frequently individuals and groups switch between the investment alternatives RISK_{OWN}, $\text{RISK}_{\text{EXPERT}}$, and RISKFREE over the course of the experiment. Table 3 outlines average and median switching frequencies for treatments as well as average switching frequencies conditional on gender.[21] Across treatments we find a strong group effect as switching frequencies in Treatment GROUP are significantly lower compared to the other treatments. This is likely attributable to the higher coordination effort necessary in this treatment. Separating results by gender it is evident that men switch less frequently between investment alternatives than women in Treatment INDIV. Gender differences in treatments COMM and GROUP point into the same direction but are insignificant.

Table 3: Average switching frequencies between investment alternatives RISK_{OWN}, $\text{RISK}_{\text{EXPERT}}$, and RISKFREE (median values in parentheses). ALL comprises the full sample, M (F) denotes male (female) individuals, MM denotes male only groups, MIX are mixed groups, and FF are female only groups.

Switching frequencies	INDIV	COMM	GROUP
ALL	9.57 (7.5)	7.40 (4.5)	5.63 (2.5)
M/MM	7.51 (4.0)	8.08 (4.5)	4.50 (1.5)
MIX		6.39 (4.0)	5.61 (3.0)
F/FF	12.87 (13.5)	9.56 (9.0)	6.78 (2.0)

To summarize, we find marked differences in the decision behavior between treatments. While communication has limited impact, group decision making leads to significantly more frequent decisions for RISK_{OWN} compared to the dominated option $\text{RISK}_{\text{EXPERT}}$. Also RISKFREE is chosen less frequently. Thus, decisions made in groups correspond more to expected value maximizing

[20] Note that the statistical tests presented in this section do not account for the fact that subjects must choose between one of the three alternatives. To corroborate the presented results we run a multinominal-probit regression, which accounts for this concern. The regression results are presented in Table A3 in Appendix A of the online supplement and support our main findings: $\text{RISK}_{\text{EXPERT}}$ is chosen significantly less in GROUP, over time, and by MIX and F/FF. The results for RISKFREE are supported as well: significantly lower selection probability in COMM and GROUP but significantly higher probability for MIX and F/FF. Therefore the different evaluation methods used yield identical results. We thank an anonymous referee for suggesting this analysis.

[21] Table A2 in Appendix A of the online supplement presents p-values of Mann-Whitney U-tests on treatment and gender effects.

behavior. These results support findings regarding the positive impact of group decision making on decision quality (Charness and Sutter, 2012; Kugler et al., 2012). In addition, gender differences emerge within each treatment and are especially pronounced in the RISKFREE option, which is chosen more frequently by females.

4.2 Behavioral biases

We now turn to RQ 2 on the potential effects of communication and group decision making on the hot hand fallacy and the gambler's fallacy. Remember that people prone to the hot hand fallacy (gambler's fallacy) expect a non-autocorrelated random sequence to exhibit positive (negative) autocorrelation.[22] To document biases in subjects'/groups' behavior we show their decision behav-

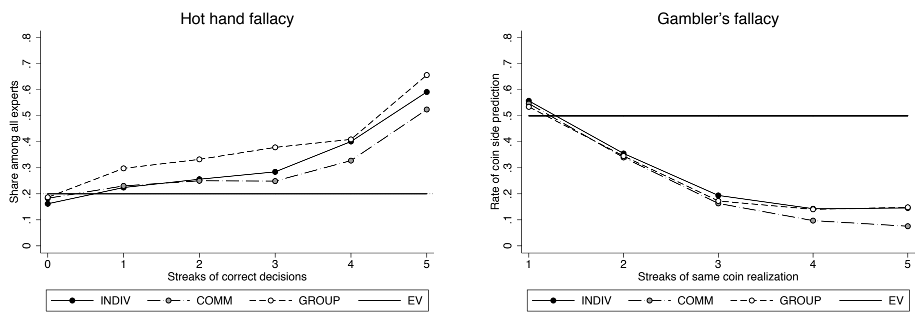


Figure 3: *Left panel*: Evidence on the hot hand fallacy by treatments measuring an expert's share among all expert decisions conditional on streaks of correct guesses. *Right panel*: Evidence for the gambler's fallacy by treatments measuring the ratio for head (tail) conditional on streaks of head (tail) realizations in the past. *EV* indicates the naïve expected share assuming unbiased decision behavior (0.2 (0.5) for $\text{RISK}_{\text{EXPERT}}$ (RISK_{OWN})).

ior conditional on the occurrence of streaks in Figure 3, i.e., either streaks of identical coin realizations in RISK_{OWN} or streaks of successful expert decisions in $\text{RISK}_{\text{EXPERT}}$. In the left panel we plot the average share of decisions an expert gains among all $\text{RISK}_{\text{EXPERT}}$ decisions conditional on his recent streak of correct decisions. Assuming unbiased decision behavior, each expert would on average gain one fifth of all decisions delegated to experts irrespective of past performance. However, what we observe is a pattern of biased behavior in each treatment. An expert's share among all expert decisions increases steadily with the number of correct decisions in the past, resulting in numbers well above the naïve expectation. This result is in line with Rabin (2002) who postulates that a subject who is affected by the overinference bias believes that a fund

[22]Note that the two are not symmetric concepts, though.

manager who is successful in two consecutive periods must be unusually good. Furthermore, these results support empirical findings in Sirri and Tufano (1998) showing that successful fund performance in the past leads to a disproportionate inflow of new investors and capital. Comparing across treatments we find no statistical differences in expert shares on the streak level (Mann-Whitney U-tests, p-values>0.10) indicating that neither communication nor group decision making influences the hot hand fallacy.[23] Thus, the overinference bias seems to map individual and group behavior quite accurately for those subjects who choose $\mathrm{RISK}_{\mathrm{EXPERT}}$.

In the right panel of Figure 3 we plot the average frequency (among all $\mathrm{RISK}_{\mathrm{OWN}}$ decisions) of choosing head (tail) conditional on streaks of head/tail realizations drawn immediately before.[24] Assuming unbiased decision behavior, each side of the coin should on average gain half of all $\mathrm{RISK}_{\mathrm{OWN}}$ decisions irrespective of past realizations. The figure reveals evidence for the gambler's fallacy as a specific side of the coin is chosen less frequently after this side exhibited a streak of several identical realizations.[25] The bias is observed in all treatments, revealing that subjects in groups are equally exposed to exhibit the gambler's fallacy compared to individuals (Mann-Whitney U-tests, $p>0.10$). These results on the gambler's fallacy are in line with the findings of Rapoport and Budescu (1997), expanded by Rabin (2002).

Our results on the appearance of both biases support findings of Ayton and Fischer (2004) who argue that people's prior expectations affect their behavior when facing random sequences in different contexts. People believe that basketball players are getting "hot" (Gilovich et al., 1985) but are less likely to develop the same belief in roulette playing. So, the hot hand fallacy is usually attributed to human skilled performance, whereas the gambler's fallacy is often observed with chance mechanisms.

[23] We use Mann-Whitney U-tests to determine statistical significance throughout this section. Test statistics and details on the test procedure are available from the authors upon request.

[24] In the right panel a streak of "0" is missing due to the fact that this would by definition equal a streak of "1" for the other coin realization yielding identical observations for streak length 0 and 1.

[25] Note that this behavior cannot be termed a "fallacy" in the strict sense, as there are no negative monetary consequences associated with it. We thank an anonymous reviewer for pointing this out.

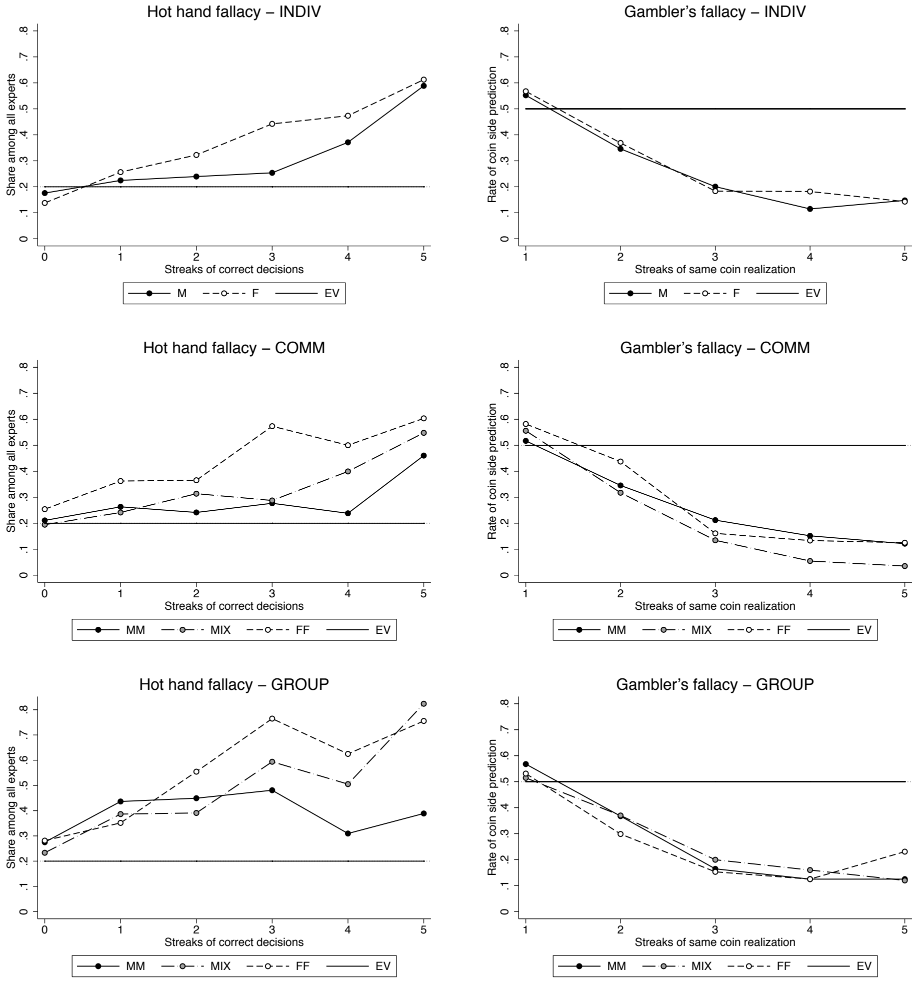


Figure 4: *Left panels*: Evidence on the hot hand fallacy by gender in each treatment measuring an expert's share among all expert decisions conditional on streaks of correct guesses. M stands for male, F for female, MM indicate male-only-, FF female-only- and MIX stand for mixed-groups. *Right panels*: Evidence for the gambler's fallacy between gender in each treatment measuring the ratio for head (tail) conditional on streaks of head (tail) realizations in the past. *EV* indicates the naïve expected share assuming unbiased decision behavior (0.2 (0.5) for $\text{RISK}_{\text{EXPERT}}$ (RISK_{OWN})).

In Figure 4 we deepen the analysis by splitting the sample by gender. The left (right) panel repeats the analysis for $\text{RISK}_{\text{EXPERT}}$ (RISK_{OWN}). In line with the visual impression of the graphs we report a weak gender effect, indicating that women (INDIV) and female-only groups (COMM and GROUP) show a marginally higher proneness to the hot hand fallacy. The right hand side panels of Figure 4 reveal no evidence of gender effects within treatments indicating that men and women exhibit the same proneness to the gambler's fallacy.

To summarize, communication and group decision making do not cure subjects from the hot hand fallacy (overinference bias) or the gambler's fallacy. In addition, women (INDIV) and female-only groups (COMM and GROUP) show a marginally higher proneness to the hot hand fallacy. These findings indicate limits to the superior performance of groups compared to individual decision making.

5 Conclusion

We reported results from decision experiments where subjects predicted coin tosses themselves, delegated the decision to experts or chose a risk-free alternative. We analyzed three treatments which were distinguished by the role of communication and group decision making: In the benchmark treatment INDIV decisions were made individually. In treatments COMM and GROUP subjects were assigned to groups of two and a chat was installed. While communication was possible in both treatments, they differed in the way decision making took place. In Treatment COMM subjects were able to communicate, but decided individually. In Treatment GROUP subjects had to agree on a decision as a group.

Subjects' decisions differed significantly across treatments. Most importantly, we showed that (i) communication and group decision making did not impact subjects' overall proneness to behavioral biases like gambler's fallacy and hot hand fallacy. (ii) Furthermore, groups in Treatment GROUP rely less on useless expert advice compared to the other treatments. (iii) Group decision making in Treatment GROUP led to fewer choices of the risk-free alternative and to more own guesses on the realization of the coin toss compared to the other treatments. (iv) Finally, we observed that gender composition of groups played a crucial role in investment behavior: groups of two female subjects choose the risk-free investment significantly more often and delegated investment decisions less often to experts than groups of two male subjects. In addition, we are the first to document that women (INDIV) and female-only groups (COMM and GROUP) showed a marginally higher proneness to the hot hand fallacy.

The main contribution of this paper is twofold. The first novel contribution is the finding that groups do not overcome hot hand fallacy and gambler's fallacy. This result is remarkable and deserves further investigation as it contrasts literature showing the superiority of groups compared to individuals. However, this superiority of groups holds in strategic (e.g., Feri et al. 2010; Sheremeta and Zhang 2010; Cheung and Coleman 2011; Sutter et al. 2013) and non-strategic situations (Blinder and Morgan 2005; Charness et al. 2007; Sutter 2007; Charness and Sutter 2012). In addition, the second major contribution shows that groups act more according to a risk-neutral benchmark (maximizing expected value) corroborating findings in Kugler et al. (2012). Groups invest in the risky investment more frequently, choose the risk-free alternative less often and rely less on outside advice compared to individuals.

References

Ackert, Lucy F., Bryan K. Church, Li Qi. 2012. An experimental examination of portfolio choice. Working Paper.

Asparouhova, Elena, Michael Hertzel, Michael Lemmon. 2009. Inference from streaks in random outcomes: Experimental evidence on beliefs in regime shifting and the law of small numbers. *Management Science* **55**(11) 1766–1782.

Ayton, Peter, Ilan Fischer. 2004. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition* **32**(8) 1369–1378.

Barber, Brad M., Terrance Odean, Lu Zheng. 2005. Out of sight, out of mind: The effects of expenses on mutual fund flows. *Journal of Business* **78**(6) 2095–2120.

Blinder, Alan S., John Morgan. 2005. Are two heads better than one? Monetary policy by committee. *Journal of Money, Credit, and Banking* **37**(5) 789–811.

Bone, John, John Hey, John Suckling. 1999. Are groups more (or less) consistent than individuals? *Journal of Risk and Uncertainty* **18**(1) 63–81.

Bär, Michaela, Alexander Kempf, Stefan Ruenzi. 2011. Is a team different from the sum of its parts? Evidence from mutual fund managers. *Review of Finance* **15**(2) 359–396.

Brown, Keith C., W.V. Harlow, Laura T. Starks. 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance* **51**(1) 85–110.

Carhart, Mark M. 1997. On persistence in mutual fund performance. *The Journal of Finance* **52**(1) 57–82.

Casari, Marco, Jingjing Zhang, Christine Jackson. 2012. When do groups perform better than individuals? A company takeover experiment. doi:http://dx.doi.org/10.2139/ssrn.1673173. URL `http://ssrn.com/abstract=1673173`. University of Zurich Working Paper No. 504.

Charness, Gary, Edi Karni, Dan Levin. 2007. Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty* **35**(2) 129–148.

Charness, Gary, Edi Karni, Dan Levin. 2010. On the conjunction fallacy in probability judgement: New experimental evidence regarding linda. *Games and Economic Behavior* **68**(2) 551–556.

Charness, Gary, Matthias Sutter. 2012. Group decision-making: More rational and less behavioral? *Journal of Economic Perspectives* **26**(3) 157–176.

Chen, Gongmeng, Kenneth A. Kim, John R. Nofsinger, Oliver M. Rui. 2007. Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making* **20**(4) 425–451.

Cheung, Stephen L., Andrew Coleman. 2011. League-table incentives and price bubbles in experimental asset markets. Working Paper (IZA DP No. 5704).

Chevalier, Judith, Glenn Ellsion. 1997. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* **105**(6) 1167–1200.

Croson, Rachel, Uri Gneezy. 2009. Gender differences in preferences. *Journal of Economic Literature* **47**(2) 1–27.

Dohmen, Thomas, Armin Falk, David Huffman, Felix Marklein, Uwe Sunde. 2009. Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior & Organisation* **72**(3) 903–915.

Fahr, René, Bernd Irlenbusch. 2011. Who follows the crowd–groups or individuals? *Journal of Economic Behavior & Organisation* **80**(1) 200–209.

Feri, Francesco, Bernd Irlenbusch, Matthias Sutter. 2010. Efficiency gains from team-based coordination – large-scale experimental evidence. *American Economic Review* **100**(4) 1892–1912.

Fischbacher, Urs. 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* **10**(2) 171–178.

Galbo-Jørgensen, Claus B., Sigrid Suetens, Jean-Robert Tyran. 2013. Predicting lotto numbers: A natural experiment on the gambler's fallacy and the hot hand fallacy. Working Paper.

Gilovich, Thomas, Robert Vallone, Amos Tversky. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* **17**(3) 295–314.

Goetzmann, William N., Alok Kumar. 2008. Equity portfolio diversification. *Review of Finance* **12**(3) 433–463.

Greiner, Ben. 2004. *Forschung und wissenschaftliches Rechnen 2003, An Online Recruitment System for Economic Experiments*. In Forschung und wissenschaftliches Rechnen 2003, ed. Kurt Kremer and Volker Macho, 79-93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen., 79–93.

Huber, Jürgen, Michael Kirchler, Thomas Stöckl. 2010. The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and Decision* **68**(4) 445–462.

Keck, Steffen, Enrico Diecidue, David Budescu. 2014. Group decisions under ambiguity: Convergence to neutrality. *Journal of Economic Behavior & Organisation* **103** 60–71.

Kroll, Yoram, Haim Levy, Amnon Rapoport. 1988. Experimental tests of the mean-variance model for portfolio selection. *Organizational Behavior and Human Decision Processes* **42**(3) 388–410.

Kugler, Tamar, Edgar E. Kausel, Martin G. Kocher. 2012. Are groups more rational than individuals? A review of interactive decision making in groups. *WIREs Cognitive Science* **3**(4) 471–482.

Malkiel, Burton G. 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives* **17**(1) 59–82.

Malkiel, Burton G. 2005. Reflections on the efficient market hypothesis: 30 years later. *Financial Review* **40**(1) 1–9.

Masclet, David, Nathalie Colombier, Laurent Denant-Boemont, Youenn Loheac. 2009. Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organisation* **70**(3) 470–484.

Odean, Terrance. 1998. Are investors reluctant to realize their losses? *The Journal of Finance* **53**(3) 1175–1789.

Powdthavee, Nattavudh, Yohanes E. Riyanto. 2014. Would you pay for transparently useless advice? A test of boundaries of beliefs in the folly of predictions. *Review of Economics and Statistics* **forthcoming**.

Rabin, Matthew. 2002. Inference by believers in the law of small numbers. *The Quarterly Journal of Economics* **117**(3) 775–816.

Rabin, Matthew, Dimitri Vayanos. 2010. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies* **77**(2) 730–778.

Rapoport, Amnon, David V. Budescu. 1997. Randomization in individual choice behavior. *Psychological Review* **104**(3) 603–617.

Rockenbach, Bettina, Abdolkarim Sadrieh, Barbara Mathauschek. 2007. Teams take the better risks. *Journal of Economic Behavior & Organization* **63**(3) 412–422.

Shapira, Zur, Itzhak Venezia. 2001. Patterns of behavior of professionally managed and independent investors. *Journal of Banking and Finance* **25**(8) 1573–1587.

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

Sirri, Erik R., Peter Tufano. 1998. Costly search and mutual fund flows. *Journal of Finance* **53**(5) 1589–1622.

Suetens, Sigrid, Jean-Robert Tyran. 2012. The gambler's fallacy and gender. *Journal of Economic Behavior & Organisation* **83**(1) 118–124.

Sutter, Matthias. 2007. Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. *Economics Letters* **97**(2) 128–132.

Sutter, Matthias. 2009. Individual behavior and group membership: Comment. *American Economic Review* **99 (5)** 2247–2257.

Sutter, Matthias, Simon Czermak, Francesco Feri. 2013. Strategic sophistication of individuals and teams in experimental normal-form games. *European Economic Review* **64** 395–410.

Tversky, Amos, Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological Bulletin* **76**(2) 105–110.

Weber, Martin, Colin F. Camerer. 1998. The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization* **33**(2) 167–184.

Hot Hand and Gambler's Fallacy in Teams: Evidence from Investment Experiments*

Thomas Stöckl† Jürgen Huber‡
Michael Kirchler§ Florian Lindner¶

Online Supplement

*Financial Support from OeNB grant # 14953 is gratefully acknowledged.
†Corresponding author. Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria e-mail: thomas.stoeckl@uibk.ac.at.
‡Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria.
§Innsbruck University, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria and Centre for Finance, Department of Economics, University of Gothenburg, Vasagatan 1, 41124 Gothenburg, Sweden.
¶Innsbruck University, Department of Public Finance, Universitätsstrasse 15, 6020 Innsbruck, Austria.

Appendix A: Additional Tables

Table A1: Coin realization and expert performance.

Period	**Coin**	**Expert 1**	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**
1	Tail	L	W	W	L	L
2	Head	W	W	W	L	W
3	Tail	W	L	W	L	L
4	Head	W	L	L	L	W
5	Tail	L	L	L	L	L
6	Tail	L	W	L	W	W
7	Tail	W	W	W	W	L
8	Head	W	L	L	L	L
9	Head	W	L	L	L	L
10	Head	L	W	W	W	L
11	Tail	L	L	W	L	L
12	Head	L	W	W	W	W
13	Tail	L	L	W	L	L
14	Head	W	L	W	L	L
15	Tail	W	W	L	L	L
16	Tail	W	W	W	L	W
17	Tail	L	L	W	L	L
18	Head	L	L	W	W	L
19	Tail	W	L	L	W	W
20	Tail	L	W	L	W	W
21	Head	L	W	W	L	L
22	Tail	W	L	W	L	L
23	Tail	W	W	L	L	L
24	Head	W	L	L	W	W
25	Head	L	L	W	W	W
26	Head	L	W	L	L	W
27	Tail	L	W	L	W	W
28	Head	W	W	W	L	L
29	Head	L	L	L	L	L
30	Tail	W	W	W	W	L
31	Tail	L	L	W	W	W
32	Head	W	L	L	L	L
33	Head	L	W	W	W	W
34	Head	L	W	L	W	W
35	Head	L	W	W	L	W
36	Head	L	L	L	W	L
37	Tail	L	W	W	W	L
38	Head	W	W	L	L	L
39	Tail	L	W	L	L	L
40	Tail	W	L	L	L	L
$\sum$ Tail/W	20	18	21	21	16	15
$\sum$ Head/L	20	22	19	19	24	25

Notes: Experts: L=lose; W=win.

Table A2: P-values from Mann-Whitney U-tests of testing for differences in switching frequencies between investment alternatives RISK_{OWN}, $\text{RISK}_{\text{EXPERT}}$, and RISKFREE. *Top panel*: treatment comparisons. *Bottom panel*: gender comparison. M (F) denotes male (female) individuals, MM denotes male only groups, MIX are mixed groups, and FF are female only groups.

Treatment comparison	INDIV-COMM	INDIV-GROUP	COMM-GROUP				
	0.2951	0.0088	0.0144				
Gender comparison	INDIV	COMM			GROUP		
	M-F	MM-FF	MM-MIX	MIX-FF	MM-FF	MM-MIX	MIX-FF
	0.002	0.271	0.287	0.070	0.496	0.052	0.673

Table A3: Multinomial probit regressions on the investment alternatives with standard errors (in parentheses), clustered at the individual level for INDIV and group level for COMM/GROUP. ***, ** and * represent significance at the 1, 5 and 10 percent levels, respectively.

Decisions	RISK_{OWN}	$\text{RISK}_{\text{EXPERT}}$	RISKFREE
α		−0.168	−1.991∗∗∗
	B	(0.145)	(0.198)
COMM	A	−0.030	−0.334∗
	S	(0.170)	(0.173)
GROUP	E	−0.378∗∗	−0.684∗∗∗
	L	(0.179)	(0.241)
Period	I	−0.032∗∗∗	0.002
	N	(0.003)	(0.003)
Group Comp.	E	−0.151∗	0.298∗∗∗
		(0.090)	(0.100)
N	14.400		
Clusters	240		
Prob > χ^2	0.000		

Notes: COMM and GROUP are treatment dummies; *Period* is a period indicator and runs from 1 to 40; *Group Comp.* distinguishes group composition 0=M/MM, 1=MIX, 2=F/FF.

Table A4: Subjects demographics by treatments. Standard deviations are in parentheses, and p-values are from Kruskal-Wallis-tests.

Treatments	INDIV	COMM	GROUP	***p-values***
Female (%)	38.3 (4.46)	41.7 (4.52)	50.0 (4.58)	0.2735
Age	22.9 (0.25)	23.1 (0.24)	23.6 (0.42)	0.3835
Semester	6.6 (0.31)	6.1 (0.34)	6.5 (0.31)	0.3041
Overconfidence	2.1 (0.06)	2.2 (0.06)	2.3 (0.06)	0.1473
Stock market experience	1.7 (0.04)	1.7 (0.04)	1.8 (0.04)	0.3423
Mood	0.9 (0.06)	1.1 (0.07)	1.1 (0.08)	0.3356

Notes: *Female* represents the percentage share of participating female subjects; *Age* is the subjects' age in years; *Semester* is the average number of semesters; *Overconfidence* is the average value to the question "Do you think that your final payment is above the average of all participants?" (1 = "above average"; 2 = "average"; 3 = "below average"); *Stock market experience* is the average value to the question "Do you have investment experience (e.g. did you buy stocks)?" (0 = "no"; 1 ="yes"); *Mood* is the average value to the question "How did you feel after the experiment" (1 = "excellent"; 2 = "good"; 3 = "rather poor": 4 = "poor").

Table A5: Subjects demographics by sessions 1 to 6 (S1 – S6) in Treatments INDIV, COMM, and GROUP. P-values are from Kruskal-Wallis-tests (standard deviations available upon request). ***, ** and * represent significance at the 1, 5 and 10 percent levels, respectively.

INDIV	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	***p-values***
Female (%)	50.0	35.0	30.0	35.0	40.0	40.0	0.9256
Age	22.2	22.7	23.6	23.7	23.3	22.4	0.0895*
Semester	5.7	7.8	7.0	7.0	6.1	6.0	0.2116
Overconfidence	1.9	2.2	2.3	2.1	2.3	2.1	0.3957
Stock market experience	1.8	1.8	1.7	1.8	1.8	1.6	0.8907
Mood	0.5	1.0	1.2	1.1	1.0	0.9	0.0579*
COMM	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	***p-values***
Female (%)	30.0	55.0	35.0	60.0	35.0	35.0	0.4586
Age	23.4	23.7	23.8	22.4	23.0	22.7	0.6130
Semester	7.6	5.7	6.3	5.0	6.5	5.7	0.3313
Overconfidence	1.9	2.1	2.2	2.2	2.5	2.2	0.3278
Stock market experience	1.7	1.8	1.6	1.6	1.7	1.7	0.8054
Mood	0.7	1.2	1.1	0.9	1.6	1.4	0.0541*
GROUP	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	***p-values***
Female (%)	75.0	60.0	45.0	25.0	50.0	45.0	0.1389
Age	22.4	23.8	23.0	25.8	23.0	23.6	0.1659
Semester	6.0	6.5	6.8	7.3	6.2	6.0	0.8527
Overconfidence	2.1	2.6	2.4	2.2	2.2	2.4	0.1953
Stock market experience	1.9	1.9	1.8	1.6	1.8	1.8	0.5832
Mood	1.1	1.3	1.0	0.9	1.1	1.2	0.6214

Notes: *Female* represents the percentage share of participating female subjects; *Age* is the subjects' age in years; *Semester* is the average number of semesters; *Overconfidence* is the average value to the question "Do you think that your final payment is above the average of all participants?" (1 = "above average"; 2 = "average"; 3 = "below average"); *Stock market experience* is the average value to the question "Do you have investment experience (e.g. did you buy stocks)?" (0 = "no"; 1 ="yes"); *Mood* is the average value to the question "How did you feel after the experiment" (1 = "excellent"; 2 = "good"; 3 = "rather poor": 4 = "poor").

Appendix B: Experimental Instructions for Treatment COMM and GROUP[1]

Dear Participant! We welcome you to this experimental session and kindly ask you to refrain from talking to each other for the duration of the experiment. If you face any difficulties, contact one of the supervisors.

Background of the experiment

This experiment is concerned with replicating investment decisions that are made on asset markets. For the case of the experiment the decisions are simplified. During the experiment you may communicate with an (anonymous) partner via a chat device to exchange your knowledge and experience. *You make your decisions together with an (anonymous) partner via a chat device.*

Market Procedure

Each participant receives an initial endowment of 500 Taler, the experiment currency. Gains and losses that are made during the experiment are added or subtracted from your current holdings. At the end of the session your Taler holdings will be converted into euros at a 100:1 exchange rate. All gains and losses are your own and payment at the end will take place privately. The experiment lasts for 40 periods.

During the course of the experiment you make investment decisions *(together with your partner).* Your chat partner is determined at the beginning of the experiment and remains the same for all periods. Because the decision making process is simplified your possibilities are limited to two different investments. You can go for an investment that involves a certain amount of risk or you can pick a risk free alternative. Taking the risky investment you have to choose between two possible outcomes. In the experiment they are named “Head” and “Tail”. Like when you toss coins, only one of the two sides can be on top and decides the game. The probability for each side is therefore 50 percent.

Beside these possibilities, there are five “experts” in the market who claim that they are able to predict the market development (the coin) better than the majority of all market participants. They always invest in the risky alternative. Like on the real market these experts offer their knowledge to everyone who is interested in it and take over the investment decision for their customers. For this service they charge an issue surcharge and a management fee. If you hand over your decision to one of the experts he will opt for either “Head” or “Tail”. At the end of each period a random process determines, which side of the coin is on top: either “Head” or “Tail”. This result and your investment decision

[1] Instructions and sceenshots are for COMM, text changes in GROUP are in (*italic*). Instructions for INDIV can be found in Huber et al. (2010).

influence your final payment.

Total Wealth (for each participant)
Each decision you take will change your current Taler holdings. If your investment decision (which side of the coin will be on top) is correct 100 Taler will be added to your holdings. If you are incorrect 50 Taler will be subtracted. If you take the risk free possibility 10 Taler will be added no matter, which side of the coin is presented. *If you and your chat partner cannot agree on a common decision, a penalty of 50 Taler will be subtracted from your account, independent of your actual decision.*
These figures change if you hand over your decision to an expert. The conditions are comparable to conditions charged by investment funds on the real market. For the access to their knowledge they claim an issue surcharge of 5 Taler, which has to be paid only for the first period you follow a certain expert. If you trust the same expert again in the directly following periods this amount will not be charged again. Experts demand a second fee. This management fee has to be paid in each period you hand over your decision to an expert and will amount 1 Taler. If the expert's decision is correct 100 Taler minus the payable fees will be added to your Taler holdings. If the decision is incorrect 50 Taler plus the payable fees will be subtracted from your account.

e.g.:	own decision:	cor.: + 100 Taler	incor.: - 50 Taler
	expert X (first time):	cor.: + 100 - 5 - 1 = + 94 Taler	incor.: -50-5-1 = - 56 Taler
	expert X (con. periods):	cor.: + 100 - 1 = + 99 Taler	incor.: - 50 - 1 = - 51 Taler

At the end of the experiment a history screen informs you about all your decisions and the outcomes. After that, your payment will be calculated according to the following schema.

100 Taler = 1 Euro

How to make your decision
You will now be informed on how to make your decisions. For a better understanding please see the screen-shot of the chat screen and the decision screen on the following pages. All relevant elements and details will now be explained.
Each period consists of the following steps:

- Chat with your partner - early leaving or end of time (1.5 min, reduced to 1 min from period 16 on).
- Individual input of your decision.
- *If both decisions are identical, the corresponding payoffs are realized (see below) and you proceed to the next period.*

- *If decisions are different, you will be redirected to the chat for 45 sec. to reach a common decision. If decisions deviate a second time, 50 Taler are subtracted from your account.*

Chat screen

In the upper part of the chat screen information about the number of periods and the remaining time for the current decision are displayed. Below this information you can see your member number within your group (M1 or M2) and your current Taler holdings.

In the middle of the screen a history screen is displayed. Here you find information about the number of periods played, your investment decisions, which side of the coin was on top, how that effected your result and how your holdings changed each period. The following 5 columns inform you about the performance of the 5 experts. If their investment decision was correct this is signaled via "!!!" (three exclamation marks). An incorrect decision is signaled via "—" (level bar). Finally the row below the history box provides you some additional information about the performance of the experts. The number stands for a performance ratio: it tells you how many of the previous 5 periods the experts´ decisions were correct (0.80 means that the expert was correct 4 out of 5 periods).

The chat can be found in the lower part of the screen. You have 1.5 min (1 min after period 16) to communicate with your partner and reach a common decision but the entry of your decision in the decision screen is made individually. The button in the upper right part of the chat allows you to leave the chat early.

Decision screen

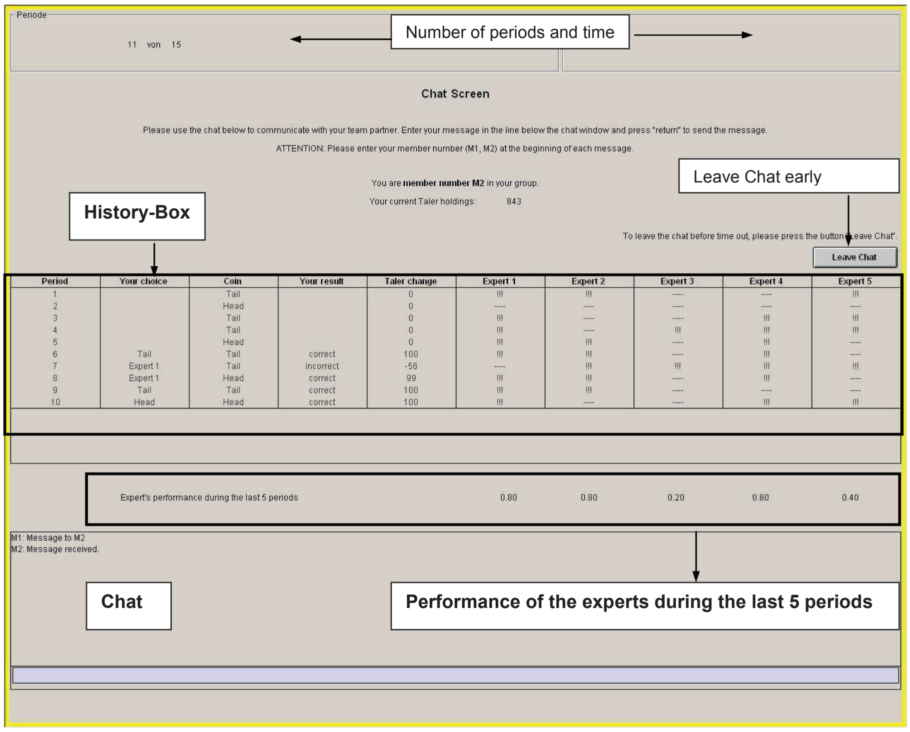
If the chat is finished (either through time out or because both group members determined it) you will enter the decision stage. The chat disappears, but the history screen remains unchanged. Above the history, there are the 8 investment buttons. Via pushing one of them you make your decision for the current period. If you like to decide on your own the buttons "Head", "Tail" or "risk free" are relevant. To trust one of the experts push one of the 5 expert buttons. *You earn a positive payment only if you and your partner agree on a common decision. If your decisions differ from each other, you are redirected to the chat for another 45 sec. Then each group member enters the decision individually. If you were not able to agree on a common decision, 50 Taler are subtracted from each of the group members' accounts.*

Important details:

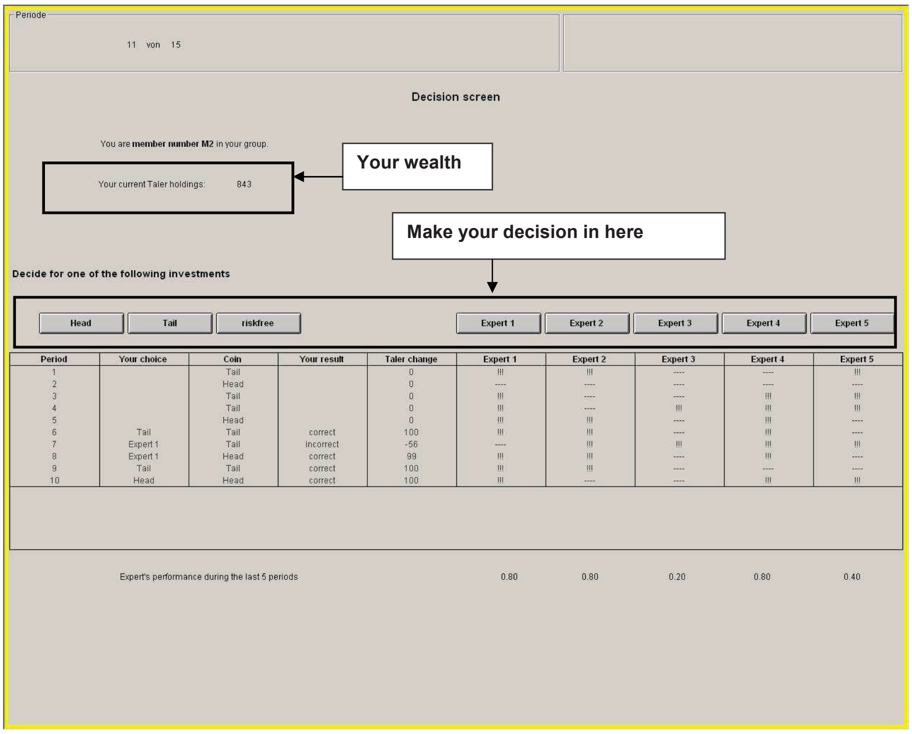
- As soon as you made your decision via a click on the corresponding button, it is not possible to undo or change your decision.

- To give you an impression of the predicting abilities of the different experts the experiments starts in period 6 and will end after period 45.
- Initial endowment: 500 Taler
- Correct decision: + 100 Taler
- Incorrect decision: - 50 Taler
- *No agreement on decision: - 50 Taler*
- "Risk free": + 10 Taler
- Issue surcharge: 5 Taler (first period you choose a certain expert)
- Management fee: 1 Taler (per period)

Chat screen


Periode
11 von 15
Number of periods and time
Chat Screen
Please use the chat below to communicate with your team partner. Enter your message in the line below the chat window and press "return" to send the message.
ATTENTION: Please enter your member number (M1, M2) at the beginning of each message.
You are member number M2 in your group.
Your current Taler holdings: 843
Leave Chat early
History-Box
To leave the chat before time out, please press the button "Leave Chat".
Leave Chat
Period Your choice Coin Your result Taler change Expert 1 Expert 2 Expert 3 Expert 4 Expert 5
1 Tail 0 !!! !!! ---- ---- !!!
2 Head 0 ---- ---- ---- ---- ----
3 Tail 0 !!! ---- ---- !!! !!!
4 Tail 0 !!! ---- !!! !!! !!!
5 Head 0 !!! !!! ---- !!! ----
6 Tail Tail correct 100 !!! !!! ---- !!! ----
7 Expert 1 Tail incorrect -56 ---- !!! !!! !!! !!!
8 Expert 1 Head correct 99 !!! !!! ---- !!! ----
9 Tail Tail correct 100 !!! !!! ---- ---- ----
10 Head Head correct 100 !!! ---- ---- !!! !!!
Expert's performance during the last 5 periods 0.80 0.80 0.20 0.80 0.40
M1: Message to M2
M2: Message received.
Chat
Performance of the experts during the last 5 periods

Decision screen



References

Huber, Jürgen, Michael Kirchler, Thomas Stöckl. 2010. The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and Decision* **68**(4) 445–462.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2013-04 **Thomas Stöckl, Jürgen Huber, Michael Kirchler, Florian Lindner:** Hot hand belief and gambler's fallacy in teams: Evidence from investment experiments

2013-03 **Wolfgang Luhan, Johann Scharler:** Monetary policy, inflation illusion and the Taylor principle: An experimental study

2013-02 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Tensions between the resource damage and the private benefits of appropriation in the commons

2013-01 **Jakob W. Messner, Achim Zeileis, Jochen Broecker, Georg J. Mayr:** Improved probabilistic wind power forecasts with an inverse power curve transformation and censored regression

2012-27 **Achim Zeileis, Nikolaus Umlauf, Friedrich Leisch:** Flexible generation of e-learning exams in R: Moodle quizzes, OLAT assessments, and beyond

2012-26 **Francisco Campos-Ortiz, Louis Putterman, T.K. Ahn, Loukas Balafoutas, Mongoljin Batsaikhan, Matthias Sutter:** Security of property as a public good: Institutions, socio-political environment and experimental behavior in five countries

2012-25 **Esther Blanco, Maria Claudia Lopez, James M. Walker:** Appropriation in the commons: variations in the opportunity costs of conservation

2012-24 **Edgar C. Merkle, Jinyan Fan, Achim Zeileis:** Testing for measurement invariance with respect to an ordinal variable

2012-23 **Lukas Schrott, Martin Gächter, Engelbert Theurl:** Regional development in advanced countries: A within-country application of the Human Development Index for Austria

2012-22 **Glenn Dutcher, Krista Jabs Saral:** Does team telecommuting affect productivity? An experiment

2012-21 **Thomas Windberger, Jesus Crespo Cuaresma, Janette Walde:** Dirty floating and monetary independence in Central and Eastern Europe - The role of structural breaks

2012-20 **Martin Wagner, Achim Zeileis:** Heterogeneity of regional growth in the European Union

2012-19 **Natalia Montinari, Antonio Nicolo, Regine Oexl:** Mediocrity and induced reciprocity

2012-18 **Esther Blanco, Javier Lozano:** Evolutionary success and failure of wildlife conservancy programs

2012-17 **Ronald Peeters, Marc Vorsatz, Markus Walzl:** Beliefs and truth-telling: A laboratory experiment

2012-16 **Alexander Sebald, Markus Walzl:** Optimal contracts based on subjective evaluations and reciprocity

2012-15 **Alexander Sebald, Markus Walzl:** Subjective performance evaluations and reciprocity in principal-agent relations

2012-14 **Elisabeth Christen:** Time zones matter: The impact of distance and time zones on services trade

2012-13 **Elisabeth Christen, Joseph Francois, Bernard Hoekman:** CGE modeling of market access in services

2012-12 **Loukas Balafoutas, Nikos Nikiforakis:** Norm enforcement in the city: A natural field experiment *forthcoming in European Economic Review*

2012-11 **Dominik Erharter:** Credence goods markets, distributional preferences and the role of institutions

2012-10 **Nikolaus Umlauf, Daniel Adler, Thomas Kneib, Stefan Lang, Achim Zeileis:** Structured additive regression models: An R interface to BayesX

2012-09 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** History repeating: Spain beats Germany in the EURO 2012 Final

2012-08 **Loukas Balafoutas, Glenn Dutcher, Florian Lindner, Dmitry Ryvkin:** To reward the best or to punish the worst? A comparison of two tournament mechanisms with heterogeneous agents

2012-07 **Stefan Lang, Nikolaus Umlauf, Peter Wechselberger, Kenneth Harttgen, Thomas Kneib:** Multilevel structured additive regression

2012-06 **Elisabeth Waldmann, Thomas Kneib, Yu Ryan Yu, Stefan Lang:** Bayesian semiparametric additive quantile regression

2012-05 **Eric Mayer, Sebastian Rueth, Johann Scharler:** Government debt, inflation dynamics and the transmission of fiscal policy shocks

2012-04 **Markus Leibrecht, Johann Scharler:** Government size and business cycle volatility; How important are credit constraints?

2012-03 **Uwe Dulleck, David Johnston, Rudolf Kerschbamer, Matthias Sutter:** The good, the bad and the naive: Do fair prices signal good types or do they induce good behaviour?

2012-02 **Martin G. Kocher, Wolfgang J. Luhan, Matthias Sutter:** Testing a forgotten aspect of Akerlof's gift exchange hypothesis: Relational contracts with individual and uniform wages

2012-01 **Loukas Balafoutas, Florian Lindner, Matthias Sutter:** Sabotage in tournaments: Evidence from a natural experiment *forthcoming in* *Kyklos*

University of Innsbruck

Working Papers in Economics and Statistics

2013-04

Thomas Stöckl, Jürgen Huber, Michael Kirchler, Florian Lindner

Hot hand belief and gambler's fallacy in teams: Evidence from investment experiments

Abstract
In laboratory experiments we explore the effects of communication and group decision making on investment behavior and on subjects' proneness to behavioral biases. Most importantly, we show that communication and group decision making does not impact subjects' overall proneness to biases like gambler's fallacy and hot hand belief. However, groups decide differently than individuals as they rely significantly less on useless outside advice from "experts" and choose the risk-free option less frequently. Finally, we document gender differences in investment behavior: groups of two female subjects choose the risk-free investment more often and are slightly more prone to the hot hand belief than groups of two male subjects.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)