Market Feedback Does not Eliminate Biases in the Perception of Independence

Sven Fischer*

26. September, 2009

In individual decision making experiments I test the robustness of biases in the perception of independence and whether they survive feedback about the median decision in groups of five. I argue that the median decision reflects the market clearing price in an equivalent interactive market setting. Despite better experimental control, known biases are confirmed but also qualified. While feedback reduces biases it does not affect behavior significantly and no learning over the 98 repetitions can be observed.

Keywords: Biased perception of randomness, gambler's fallacy, markets, experiment.

[*JEL*: C91, D03, G19]

1. Introduction

In both Psychology and Economics (with an emphasis on the former) a large set of literature explores individual perceptions of randomness. The majority of these studies find similar systematic deviations from normative behavior, and labeled or explained them as realizations of the "law of small numbers" (Tversky and Kahneman, 1971), "local representativeness bias" (Falk, 1981), "negative recency effect" (Budescu, 1987) "negative sequential bias" (Wagenaar, 1972) or, relating to the environment in which

*ELSE @ Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom; *email:* fischer@gmx.co.uk. The author gratefully acknowledges the financial support of the Economic and Social Research Council (UK) via the ESRC Centre for Economic Learning and Social Evolution (ELSE).

similar deviations are observed in the real world, "gambler's"– or "hot hand fallacy" (see, e.g. Clotfelter and Cook, 1993; Metzger, 1985).

Based on these regularities, several models were developed in order to account for such fallacies. Rapoport and Budescu (1997) for example base their model on a known limitation of the human working memory of 7 ± 2 items, which is widely accepted among psychologists.[1] They assume that within this limited length of realizations subjects try to maintain representativeness of a streak of random outcomes to the underlying process. Similarly Rabin (2002) and Rabin and Vayanos (2007) assume that subjects wrongly perceive future outcomes of an independently and identically distributed (i.i.d.) process to be correlated with outcomes in the past. In Rabin (2002) agents expect random draws without replacement within a moving band of (previous) outcomes and in Rabin and Vayanos (2007) outcomes of an i.i.d. process are perceived to be correlated with outcomes in history.

More recently, models of biased perceptions of randomness were applied to financial market problems. Barberis et al. (1998) formulate a model based on a version of the local representativeness bias which tries to explain underreaction of stock prices to news in the short term (up to 12 month) and overreaction over longer time horizons (4-6 years). Not only Barberis et al. (1998), but also Rabin (2002) and Rabin and Vayanos (2007), as well as several experimental and empirical studies see fallacies in the perceptions of randomness to be relevant for market interactions. While this is a reasonable conjecture, it is important to note that to my knowledge there are no existing studies which support this claim beyond reasonable doubts.

In this study I analyze a bias in individual perceptions of randomness which is either explicitly or implicitly included in all above definitions. It is a bias in the individual perception of independent random processes. In the following I will, thus, use the term ' bias in the perception of *independence*'. More precisely, the analysis looks at biases in the perception of processes which are easily understood and known to be independent. I argue that behavior is biased if it reflects any belief in a history dependence of processes like, e.g. the repeated flipping of a fair coin. This has an important (financial) market interpretation. Under the efficient market hypothesis all available information should be reflected in market prices. Any future changes in market prices can, thus, only be caused by unpredictable shocks with zero (un)conditional expectation, what implies that the probability of an increase or decrease in prices must be equal.

I run individual decision making laboratory experiments in order to test the robustness of such biases to better experimental control and whether one important aspect of market

[1] See Miller (1956) and as a more recent support of his results Bar-Hillel and Wagenaar (1991).

interaction, feedback via observable prices, reduces or even eliminates it. To be able to exclude interaction effects in small experimental groups subjects do not actually interact on markets. This paper is merely interested in individual biases rather than whether market outcomes are biased.

The remainder of this paper is organized as follows. In the following section I discuss the existing evidence before I turn to the question how it relates to markets in section 3. The experiment which was run with a total of 91 subjects is described and justified in section 4. The observed behavior is analyzed and described in section 5 and discussed in section 6 which concludes.

2. Evidence for Biases in the Perception of Independence

There are many empirical studies exploring and finding behavior which is indicative of fallacies in the individual perception of independence among, e.g. lottery (Clotfelter and Cook, 1993) and casino (Sundali and Croson, 2006) players or on parimutuel betting markets (Terrell, 1994; Williams, 1999).

However, many of these empirical studies explore perceptions of independence in environments which are rather exceptional for human behavior. Most lotteries and casino gambles exhibit winning prospects which theoretically should only be accepted by individuals seeking extraordinary high risk.[2] This contradicts the widely accepted conjecture that most investment and market behavior exhibits risk aversion what, among others, can be founded on evolutionary arguments (for a recent example see, e.g. Robson (2002)). Furthermore, people may have subjective reasons to question that the underlying processes are truly i.i.d.

One empirical study is the paper by Guryan and Kearney (2008) who observe a "*lucky store*" effect in three state lotteries in Texas. The authors analyze time series of lottery sales in stores throughout Texas and find a significant and lasting positive effect on ticket sales in stores which sold a winning ticket. This effect proved to be correlated to the size of the win and was more or less limited to to the particular lottery and store. While this superstitious behavior can be interpreted as an indication for biased perceptions of winning probabilities, it is unclear what this implies for investment or trading behavior.

Another example of an empirical paper on misperceptions of independence is the survey of basketball players and fans by Gilovich et al. (1985). Among others, they find that a huge majority of the surveyed agree with the following statement: A player has "*a better chance of making a shot after having just made his last two or three shots than*

[2] Most official lotteries offer a payout rate of 50%.

he does after having just missed his last two or three shots.". However, data of scores in games or penalty shootouts could not support such or similar regularities. Camerer (1989) tests whether such beliefs in hot-hands translate into betting behavior and betting outcomes. He finds that (illegal) bets on outcomes of basketball games are biased if one of the teams had a streak of several wins or losses in the past few games what is not justified by actual winning frequencies. The bias is stronger for streaks of wins than for streaks of losses. While biases are overall small, they are strong enough to render bets on teams with loosing streaks profitable enough to break even despite a bookmaker's charge of 10% on losing bets.

This result indicates that people have problems in identifying an i.i.d. processes as such, and attribute correlations where there aren't any. Camerer's finding of biased market outcomes makes this result even stronger. The fact that the market bias is just strong enough for an unbiased bettor to break even may suggest that unbiased bettors trade the effect away as much as is reasonable for them or that the effect is not very strong nor stable. However, my study is not concerned with random sequences where subjects first must identify the underlying process. Concerning Gilovich et al.'s results, more fundamental objections can be made. Sport events can be extremely emotional and 'irrational'. Also, there are examples of human activities in which human performance exhibits sequence effects (see, e.g. Gilden and Wilson, 1995, 1996).

Besides these direct empirical findings of biases either in the identification or in the prediction of random processes, there is a set of literature which attributes other anomalies or patterns to biases in the perception of independence. Some papers (e.g. Terrell, 1994; Williams, 1999) argue that the *favorite-longshot bias* in parimutuel betting markets is caused by biases in individual perceptions of randomness. However, these markets are rather inapt examples. Parimutuel betting markets have a strategic aspect which makes it difficult to conclude that biased market outcomes are due to individual misperceptions of randomness. Biased outcomes could be the result of wrong beliefs on behalf of early bidders concerning the betting behavior of others and consequently wrong deductions from published winning odds by late bidders. Furthermore, as for example Ottaviani and Sørensen (forthcoming) and Koessler et al. (2006) show, there are mechanisms related to the timing of bets which can explain biased outcomes without any biases in the perception of randomness or in the beliefs about the behavior of others.

With respect to behavior on financial markets, many authors see the (hypothesized and still contested)[3] disposition effect (Shefrin and Statman, 1985) as directly resulting from misperceptions in randomness: traders wrongly expect a mean reversion in the

[3] See, e.g. Bremer and Kato (1996).

performance of a security and, thus, tend to sell winners too early and hold losers too long (see, e.g. Weber and Camerer, 1998; Chevalier and Ellison, 1997, 1999).

While a belief in mean reversion is a convincing explanation for the disposition effect, it is not the only possible one. Weber and Camerer (1998) and Odean (1998) also explain the disposition effect using Kahneman and Tversky's 1979 prospect theory, assuming shifts in the reference point with respect to what changes in the price of a security are weighted as gains or losses.

The second and largest set of studies in support of fallacies in the perception of randomness are laboratory experiments. Most experimental studies can be found in the Psychology literature. Whilst many experiments don't give any monetary incentives, in some which do it is rather unclear what incentives the implemented monetary rewards induce. While this is a clear drawback, it does not necessarily question the relevance of their results and observed biases for the economic sphere. However, there are important aspects which make it questionable that these findings can be transferred to market environments. Experiments indicating such biases are of three kinds: production, recognition (judgement) or prediction tasks. A good overview of the psychology literature can be found in Bar-Hillel and Wagenaar (1991).

In production task experiments subjects are usually asked to produce a sequence of i.i.d. random draws. One persistent finding of such studies is that "human produced sequences have too few symmetries and long runs, too many alternations among events, and too much balancing of event frequencies over relatively short regions" (Lopes and Oden, 1987, p. 392). But as Bar-Hillel and Wagenaar (1991) already point out, systematic biases observed in such production task experiments "could be either accurate reflections of biased notions of randomness or biased reflections of accurate notions of randomness (or both)."

Further, there are studies indicating that proper incentives in production tasks considerably reduce biases. For example in Rapoport and Budescu (1992) subjects produce series of 150 binary choices in several treatments. One treatment replicates the traditional production task. In an other subjects play a symmetric matching pennies game over 150 rounds using the strategy method, i.e. they do not receive feedback between the 150 games. Rapoport and Budescu observe that in the matching pennies game where the only Nash equilibrium is mixing of strategies with equal probabilities, sequences of strategy choices exhibit much less biases than in the traditional production task.

Also feedback helps subjects to be better in producing random sequences. Neuringer (1986) gives subjects between 5 and 10 summary statistics about the sequence they just produced. This feedback includes, e.g. a summary statistic of a contingency table and the number of alterations (or reversals). Feedback helped subjects considerably to become better in generating i.i.d. sequences of binary outcomes.

The general criticism of production tasks is reflected in a series of recognition task experiments asking subjects to distinguish i.i.d. random processes from correlated ones. The most prominent example in this class of experiments is the study by Falk (1981) which confronts subjects with two sequences of random binary processes which they are told were produced by two pupils instructed by their teacher to record 30 tosses of a coin. Subjects are told that one of the pupils cheated and must identify the cheater. Most of these experiments confront subjects with one series which exhibits significant deviation from an i.i.d. fair binary process according to some measurements predefined by the experimenter and another series which, according to these measurements is resulting from the desired i.i.d. process. Kareev (1995) summarizes results of these recognition task experiments as follows: If the 'biased' process exhibits a higher alternation rate of 0.6 as opposed to the fair one of 0.5, subjects tend to identify the i.i.d. sequence to be biased. The most noteworthy aspect of these studies is that to some degree the experimenter himself falls for a representativeness bias. In most recognition task experiments the supposedly i.i.d. process itself is a selected process, deemed to be representative by the experimenter. Also, this task requires subjects to make inferences from rather limited sample sizes (although in some cases the sequences are rather long as for example in ??), which is a much more challenging task than merely having a correct understanding and representation of randomness. Other examples are Gilovich et al. (1985) and Lopes and Oden (1987).

2.1. Prediction Task Experiments

Experimental studies which come close to a market environment are prediction task experiments. Here, subjects must predict the outcome of the next step of a sequence of random draws. Still, many lack (proper) incentives.

Ayton and Fischer (2004), ask subjects to bet on the next outcome of a fair binary i.i.d. process. On aggregate they find biases which manifest themselves in a higher relative frequency of one particular prediction. However, such biases may still be the result of overall indifference. If at all it shows a minor leaning towards one of the two outcomes.

There are also many Prediction task experiments which ask subjects to predict the next outcome of a time series taken from the real world. (see, e.g., Schmalensee (1976)Dwyer Jr. et al. (1993)Williams (1999) Gardner and McKenzie (1985), or Bolle (1988)). Results can be summarized such that predictions do not conform to rational expectations but rather exhibit patterns of adaptive expectations. While it is difficult to relate their results to biases in the perception of independence, their results indicate that predictions (of relative changes) tend to be history dependent. Contrary to my study, subjects in these experiments are confronted with rather complex processes and first need to identify them.

The two experimental studies which are most relevant for this paper are Bloomfield and Hales (2002) and Asparouhova et al. (2005). Bloomfield and Hales (2002) explore perceptions of randomness in an incentivised prediction task experiment. Subjects see a sequence of eight independent and identically distributed binary outcomes, presented as a random walk where at each stage the process either goes UP or DOWN with equal step length. They must predict the outcome of the next i.i.d. draw which pays 100 with probability $1/2$ and nothing otherwise by submitting their valuation of this lottery. Although in one treatment condition they explain subjects at length that the underlying process is a 'random walk', subjects submit biased predictions.[4] Most importantly, however, the biases are correlated with the number of reversals[5] in the previous eight draws: When there were only few reversals subjects believed that the outcome of the eighth draw would be more likely to occur again in the ninth, whereas in sequences with a large number of reversals, subjects put higher probability on the opposed outcome. Thus, Bloomfield and Hales (2002) find evidence for regime-shifting beliefs as modeled by Barberis et al. (1998).

Asparouhova et al. (2005), however, raise an important objection to the design by Bloomfield and Hales (2002): Subjects are shown a selected set of 16 time series. These are rather special and unrepresentative of a random walk. More specifically, a Fisher exact test rejects the null that the set of sequences used in the experiment were drawn from the same distribution with p-value $p < 0.001$.[6] Asparouhova et al. (2005) run a new treatment in which subjects are presented with truly randomly generated sequences and are not told that the underlying process is a random walk. They observe that both, increasing number of reversals and length of the most recent streak, on average induce a belief in a reversal. However, they find weak (insignificant) evidence that subjects tend to predict a continuation rather than a reversal if the streak exceeds a certain length.

However, both studies use a payoff rule with a guaranteed profit of zero in experimental

[4] Unfortunately the full set of instructions are not available (requests made to the corresponding author were not answered). However, in the paper it says for example that in one treatment instructions stated that "...statistical models are unable to predict future outcomes from past ones and, on average, there is no upward or downward trend." (Bloomfield and Hales, 2002, , p. 403)

[5] The number of reversals gives how often an outcome was different to the one in the previous period.

[6] The preselection of i.i.d. draws raises the question of deception which the subjects may have sensed as half of the 16 random walks were mirrored images of the other eight.

currency at the correct prediction of 50. More specifically, subjects are asked to submit their valuation of a lottery ticket which pays 100 with $p = 1/2$ and zero otherwise. Whenever the submitted valuation $v > 50$ ($v < 50$), the subject buys (sells) $|p - 50|$ lottery tickets at prices $51, 52, ..., v$ $(49, 48, ..., v)$ and plays the lottery with these tickets once (as opposed to one lottery per ticket). From a behavioral standpoint this anchoring of the BDM mechanism (Becker et al., 1964) at the correct i.i.d. prediction of 50 is problematic. For one it remains unclear in the instructions what happens at $v = 50$ (basically nothing happens) and clearly this may push subjects to biased predictions which they otherwise may not have. One might object that this does not explain why deviations are biased. But the bias of the deviation may has nothing to do with the intrinsic understanding of the underlying process but rather with what the confused subject perceives to be the "correct" or "usual" thing to do. This effect may even be strengthened, as the examples given in the instructions depict heavily biased predictions.

To some extend these problems may be diluted by the use of tournament incentives.[7] However, it is unclear how Bloomfield and Hales (2002) describe them to the participants and whether they understand them. Unfortunately a full set of instructions was not available.[8]

Moreover, such tournament incentives are known to induce risk seeking behavior among experimental subjects (see the experimental literature starting with Bull et al., 1987). In the mechanism used in the experiment, risk seeking induces higher biases in reported valuations.

Strikingly, Asparouhova et al. (2005) repeat one treatment from Bloomfield and Hales (2002) but find only insignificant reactions. While this may be due to them using undergraduate students rather than MBA students, it nevertheless raises the question as to whether the mechanism yields a good proxy for intrinsic probabilities.

Besides these considerations concerning the methodology and the problems Asparouhova et al. (2005) had in replicating the results by Bloomfield and Hales (2002), this paper addresses an other important question. The experimental methods used by both give little room for learning. Feedback is probabilistic and participants can not observe other's actions. To test whether such biases are relevant in market interactions, however, it is crucial to test whether they persist in environments which replicate an important feature of markets and that gives ample opportunity to learn.

[7] The total session payoff was capped and participants were paid according to their relative performance.
[8] Requests made to the authors were not answered.

3. Market Interaction

There are several features of market interaction which may reduce or even crowd out biases. As already Camerer (1987) and Camerer (1992) point out (and show experimentally) these are one or all of the following effects: (i) cancelation of unbiased errors, (ii) prices driven by a few sophisticated traders (smart few), (iii) dying out of biased traders (evolution) and (iv) learning via observing other traders and/or market prices. Concerning biases in the perception of independence, existing evidence suggests that these are not random white noise what implies that mechanism (i) could not work. The smart few hypothesis deals with market outcomes. This paper, however, is concerned with the question whether individual and not market biases survive in a market interaction. Testing this hypothesis would require (potentially) too many traders with unlimited liquidity.[9] Similarly the evolution hypothesis is merely concerned with the efficiency of the mechanism. Thus, mechanisms (ii) and (iii) are best analyzed theoretically or empirically. They are not within the scope of this paper.

Instead, in this study we concentrate on whether learning via observing market prices reduces or even eliminates biases. The strongest experimental support for biases in the perception of independent processes comes from individual decision making experiments with no feedback from other participants. The most obvious information agents can freely observe in a market are the market prices. As a proxy for market prices in the feedback treatment we form groups of five subjects and inform each group member about the median valuation in his group. In the context of the actual experimental procedure where subjects are only allowed to trade one lottery ticket a period, the median is the actual market clearing price if we would have introduced an actual trading mechanism or a market with an auctioneer.

To my knowledge there exist no controlled experimental study which explores whether biases in the perception of independent processes as, e.g. the gambler's fallacy, persist in environments where subjects get feedback about the decisions of others.

4. Experiment

This study looks at biases in the prediction of the next step of a series of independent and identically distributed binary draws. It tests the robustness of biases observed by Bloomfield and Hales (2002) and Asparouhova et al. (2005) and whether feedback in form of the median response in the previous period suffices to reduce or eliminate them. The general experimental procedure is as follows.

[9] Alternatively one could run experimental markets with several human and one automated trader.

- Every subject is confronted with a total of 100 sequences of the outcomes of eight flips of a fair coin.
- As in Bloomfield and Hales (2002) these are represented by a graph of eight equidistant steps representing 'heads' by an upward movement and 'tails' by a downward movement. An example of such a graph is given in figure 1.
- Subjects submit their valuation v of a lottery ticket paying 100 if next draw is "UP" and zero otherwise before the final (9th) draw.
- The "price" r of the lottery ticket (terminology used in instructions) is determined by another random draw from a uniform distribution with $r \in [0, 100]$.
- If $r < v$ the subject buys one asset at r; if $r > v$ he short sells at r and if $r = v$ he neither buys nor sells.
- Outcome $l \in \{0, 100\}$ of 9th draw determines resulting payoffs:

$$\pi = \begin{cases} l - r & \text{if } r < v \\ 0 & \text{if } r = v \\ r - l & \text{if } r > v \end{cases}$$

Every subject goes through this procedure 100 times of which the first two repetitions at the beginning are unpaid trial rounds (without feedback about the median decision). With respect to the main research question I compared a feedback treatment to the above baseline treatment between subjects, i.e. every subject either played the baseline or the feedback treatment. In the feedback treatment subjects were formed into groups of five. All subjects within that group observed the same sequence. At the end of every round subjects were informed about the median valuation reported in their group of five. In order to keep experimental procedures[10] and data structure as similar as possible, subjects in the baseline treatment were also formed into groups. As in some sessions not enough subjects showed up to the experiment, these group sizes varied between four and five.

All 98 repetitions were paid. Paying a random selection of outcomes instead would clearly reduce potential endowment effects. However, this would be technically infeasible and make experimental costs rather unpredictable. Furthermore, there is always the risk that someone submits extreme valuations. For example, someone who always submits

[10] In groups subjects always need to wait until the slowest one has reached a decision.

$v = 0$ or $v = 100$ risks to make losses overall with probability 1/2. We therefore must monitor incomes and exclude those from further participation who have accumulated too many losses with respect to the show up fee. To cover for potential losses in the early stages of the experiment, subjects received a flat fee of 500 experimental currency units.[11]

[Insert Figure 1 about here.]

Before describing some aspects of the experiment in more detail, let me elaborate more on the changes in the design compared to Bloomfield and Hales (2002).

First of all, this experiment uses a different elicitation mechanism with positive expected payoff in experimental currency at the normative prediction. It does not rely on tournament incentives and is incentive compatible under risk neutrality and some forms of risk preferences. Note, however, that this is not a requirement in order to be able to draw meaningful conclusions from responses. For this we only require that a (un)biased perception is associated with a (un)biased valuation. Under expected utility the mechanism used in this experiment always elicits a biased prediction if the intrinsic probability is biased (irrespective of risk preferences) and an unbiased one if it is not. More generally, under expected utility theory it satisfies the following symmetry property.

Theorem 1 The utility maximizing valuation v^* for some subjective probability p^e and the valuation v^{**} for the alternative $1 - p^e$ are symmetric around $1/2$. More specifically, if $v^* = \tilde{v}$, then $v^{**} = 100 - \tilde{v}$.

A proof of these properties of the mechanism is given in the appendix.

An important consequence of this symmetry property is that we can test whether biased subjects suffer from an 'optimism' or 'pessimism' bias. More specifically, we can test whether the size of a bias is correlated with the general direction of the sequence. For example, assume that for a particular sequence a subject submits a valuation of $v_1 = 60$. However, for an identical but mirrored sequence he reports a valuation of $v_2 = 45$ instead of 40 which would be the equivalent downward biased response. If now, the first sequence was overall upward moving, i.e. outcome 'UP' occurred more than four times, this would indicate an 'optimism' bias.

Bloomfield and Hales (2002) and Asparouhova et al. (2005) use two different descriptions of the underlying process. In one treatment they explain the process as resulting

[11] Only one subject (who mostly submitted valuations of 0 or 100), accumulated too many losses and we had to ask her to leave. Fortunately this happened in the baseline treatment.

from the repeated flipping of a fair coin. In an other they explain that the UP and $DOWN$ movements represent typical "performance surprises" of firms.

However, representing the process as a sequence of performance surprises, at least with a normal student subject pool, may be problematic. Reactions in this frame are only meaningful if the subject has experience or knowledge about such processes. This is especially true as it is almost impossible to make meaningful inferences on the properties of the process from short sequences length 8. We, thus, only use one frame which truthfully explains the easily understood and commonly known process as a sequence of the outcomes of repeated flips of a fair coin.

In reality the sequences were actually produced by repeated flips of a fair coin. Due to time constraints, however, these were not made in front of the subjects but pre-recorded and imported into the experimental program. Similarly to Dwyer Jr. et al. (1993) the random sequences were tested for significant deviations. This was made by testing for every period whether the sequences from ten consecutive rounds[12] differ significantly ($p < 5\%$) from an i.i.d. fair binary process. This actually led to the exclusion of two recorded sets of 100 sequences.

With a total of 98 paid repetitions per subject we were able to obtain enough observations of reactions to sequences with rather rare patterns like, e.g sequences in which one outcome occurs seven times. Furthermore, this gave subjects ample opportunity to learn from probabilistic feedback, and in the feedback treatment also from the information about the median decision.

As previous studies established that biases are robust to variations in the representation of the sequence, no further representation than the illustrated graph was used or tested.

5. Results

A total of 91 subjects participated in 6 sessions in the computer laboratory of the ESRC Centre for Economic Learning and Social Evolution at University College London. Subjects were recruited via the online recruitment system in which subjects can sign in online and which guarantees that no subject participates more than once.[13] To obtain a homogenous subject pool for better identification of between subjects effects, only undergraduate students from Universities in London were invited.

[12] This is motivated by findings concerning human memory made, e.g. by Miller (1956). With nine previous periods plus the current one, a total of ten rounds covers approximately the working memory of almost everyone.

[13] The recruitment system uses the software ORSEE (Greiner, 2004).

In the baseline treatment without feedback, a total of 46 subjects divided into 10 groups participated and in the treatment with feedback a total of 45 subjects in 9 groups of five subjects each. In the baseline treatment some groups consisted of only four subjects what was due to a lack of subjects showing up for the session. Every group in both treatments confronted a different sequence of random draws which were prerecorded by the repeated flipping of a coin.

Including a show up fee of £ 5 on average subjects earned £ 14.18 with standard deviation 3.49 in about 75 and 90 minutes, including admission and payment.

Looking first at the submitted valuations, average valuations are 47.64 in the baseline treatment and 48.49 in the treatment with feedback. A Wilcoxon rank sum test on the distribution of average valuations of every subject can not reject the null hypothesis of no significant shift in location due to feedback. If not mentioned otherwise, significance levels are set to $\alpha = 2.5\%$ throughout. In the treatments with feedback, the reported median was exactly 50 in 62.43%, above 50 in 20.85% and below 50 in 16.72% of all cases. Still, the overall mean of 49.42 (with standard deviation 12.75) is close to the unbiased prediction. We can, thus, expect subjects to be able to learn from the observed median.

Signed Deviations

Unfortunately, the submitted valuations themselves are not very informative. Instead the following section analyzes signed deviations from 50 (SDV) which is defined as follows: If the 8th move on the graph was 'UP', then $SDV = v - 50$, otherwise $SDV = 50 - v$. Thus, a positive SDV indicates a belief in a repetition of the 8th outcome and a negative in a reversal.

The average SDV's are -2.075 in the baseline and -2.513 in the feedback treatment, with standard deviation (over averages of every subject) 5.272, and 5.275, respectively. Again, a Wilcoxon signed rank test does not reject that null that the local shift parameter equals zero.

Let us turn to the question how SDV's depend on the stimuli. Table 1 lists the estimates (and standard errors) for several regressions which were made separately for each treatment. All regressions are linear mixed effects models with a random shift effect per group and random effect on all regressors per subject nested in group. The main variables of interest are the length of the last streak, the number of reversals and the frequency of the 8th outcome over the entire sequence. More specifically, variable 'Streak' measures how often the 8th outcome occurred in a row at the end of the sequence, 'Reversals' measure how often two consecutive draws were different, and variable 'Freq8th' measures how often the last outcome occurred within the entire sequence. For example, sequence (UP, DOWN, DOWN, UP, DOWN, UP, UP, UP) has a streak length of three, a total of four reversals and the 8th outcome occurred five times.

As indicated by the suffixes *dm*, regressors were de-meaned by substracting the expected values which are 510/256, 3.5, and 4.5, for Streak, Reversals, and Freq8th. In addition, the regressions test for learning effects over time by inclusion of variable period. For data from the treatment with feedback, I also test whether there is immediate learning from the previously observed median.

De-meaning the regressors allows us to interpret the sign of the coefficient on the intercept as a general tendency. A negative intercept like the ones we observe in the regressions, indicates that there was a tendency to believe more in a reversal than in a continuation. Note, however, that this effect is only significant in the feedback treatment.

[Insert Table 1 about here.]

From the regression results in table 1 it is obvious that all three main variables significantly affect behavior. There are no learning effects over time in both treatments and the previously observed median also does not affect behavior. In other regressions I tested interaction effects with variable period and with a variable measuring the average observed median per group. These, however, do all not affect reported valuations.

During the analysis it became clear that reactions to these variables are not linear over the entire range of the regressors. Thus, all main regressors were transformed by splitting them into two variables, each only either including the negative ('< 0') or positive ('> 0') cases and being zero otherwise. Models (1) and (3) are models which include all main regressors of interest and models (2) and (4) are the best models resulting from a stepwise exclusion of insignificant and unexplanatory variables, tested by likelihood ratio tests. Note that model (4) is run on more data than model (3) as the exclusion of the lagged median allowed to include observations from the first paid period.

[Insert Table 2 about here.]

Lets first turn to the baseline treatment with Info=0. The estimation results show that there is no general tendency in the bias and again no learning effects can be found. For streak length of 2 or higher, there is an increasing belief in a reversal with increasing streak length. Similarly, for a number of reversals of 4 or higher, an increase in the number of reversals induces a belief in a reversal for the 9th outcome. The same holds for a high frequency of the last outcome during the series of eight draws, if this frequency is 5 or higher.

Furthermore, the coefficient on dmStreak< 0 is significant. Note that this is essentially a dummy for streak length 1 and that the regressor is negative. Thus, at a streak length of one, the negative intercept is offset, what is proven by a joint hypothesis test that $-254/256 * dm\text{Streak}_{<0}+\text{Intercept}= 0$ against the general alternative that it is unequal zero.

In the feedback treatment there is a general tendency to belief in a reversal as the intercept is negative and significant. An above average streak length and frequency of the 8th outcome, again induce a belief in a reversal and like in the baseline treatment, a streak length of 1 induces a - rather weak - belief in a continuation. Quite surprisingly, the same must be said for small numbers of reversals: The smaller the number of reversals, the stronger the belief in a continuation. A very high number of reversals, however, has no effect.

This indicates that there are substantial difference in the way subjects react to the sequences, despite the fact that no significant influence of the information itself nor a change in behavior over time could be found. Table 3 lists the results of two estimations including the combined data set of both treatments. The full model, including dummy D_I indicating the treatment with feedback and it's interactions with all relevant variables, is listed in the second and last column. The best model, obtained after a stepwise reduction minimizing the Akaike information criterion, is listed in the first column only. Quite strikingly, no treatment effect can be found. Dummy D_I and all it's interactions with the other regressors are insignificant. Note, however, that - though insignificant - on streak length and number of reversals, feedback dampens the biases.

For the combined data we observe an overall tendency to believe in a reversal rather than a continuation. The longer the length of the last streak the higher the number of reversals; and the more often the last outcome occurred in the sequence, the higher the belief that the next draw is different from the 8th.

Still there is some minor evidence that subjects are more rational in the treatment with feedback. For example, in the baseline treatment only one subject submitted a valuation of 50 throughout, whereas in the treatment with feedback 7 did.[14] Allowing for some errors, 6 compared to 9 subjects submitted a valuation of 50 in 93 rounds ($\approx 95\%$ of all 98 rounds).

[Insert Table 3 about here.]

Figure 2 gives a more detailed impression of how behavior reacts to the characteristics of the sequences. The first panel plots average signed deviations from 50 by streak length,

[14] A Fisher exact test finds that this is a weakly significant difference with $p = .0585$.

the second by number of reversals and the third by the frequency of the 8th outcome. The averages illustrated by circles and bullets are estimated by linear mixed effects regressions of the signed deviations on eight dummies, one for each of the eight cases. Estimations also included random effects on all subjects nested in group effects. The plotted points are the maximum likelihood estimates of the dummies and the whiskers illustrate a 95% confidence interval. Circles illustrate an insignificant estimate, dots a significant ($p \leq 2.5\%$).

Squares illustrate the results of similar regressions which additionally control for the signed de-meaned variables of the other characteristics. For example, the squares in the first panel of figure 2 are the estimates of the first eight coefficients ($\delta_1 \ldots \delta_8$) in the following regression model:

$$\begin{aligned} SDV_{mit} &= \delta_1 D(Streak = 1)_{mit} + \delta_1 D(Streak = 2)_{mit} + \ldots + \delta_8 D(Streak = 8)_{mit} + \\ &\quad \beta_1 dm\text{Reversals} > 0_{mit} + \beta_2 dm\text{Reversals} < 0_{mit} + \\ &\quad \beta_3 dm\text{Freq8th} > 0_{mit} + \beta_4 dm\text{Freq8th} > 0_{mit} + g_m + c_i + u_{mit} \end{aligned}$$

where g_m is a random effect on group m, c_i is a random effect on subject i nested in m, and u_{mit} is a standard white noise residual. Filled squares illustrate significant, empty squares insignificant estimates. Again, whiskers illustrate a 95% confidence interval.

The circles can, thus, be interpreted as unconditional means of the signed deviations of valuations and the squares as means conditioned on the averages of the other characteristics. Clearly the conditional means for extreme realizations are rather meaningless. For example, a streak length of 8, 0 reversals and a frequency of the last outcome of 8 all describe the same two sequences.[15] Conditioning on the average number of Reversals and Freq8th is, thus, pretty meaningless. In the following we will therefore look at the unconditional means for extrem realizations and at the conditional ones for all intermediate ones.

[Insert Figure 2 about here.]

The estimated mean reactions reflect our previous results and qualify them at the same time. Obviously, long streaks have the strongest influence and induce a belief in a reversal. This already starts at a streak length of 3. Also note that, as described above, at a streak length of 1 which occurs with probability 1/2 there is no significant bias. Turning to reversals, our previous result on regressor dmReverslas< 0 is qualified. Obviously the observed reaction only comes from a strong reaction for the two sequences

[15]This is reflected in the equivalent estimates for these cases.

without any reversals. For 1 to 3 reversals, there is no significant bias. Only with above average number of reversals (4 to 8) a small, but mostly significant negative bias can be observed. Interestingly, for a very high number of reversals, valuations are very dispersed. Finally, the estimates on the frequencies of the 8th outcome, confirm our previous result that a belief in a reversal is induced if the frequency of the last outcome was above average.

Mirrored Sequences

In the description of the mechanism we had established a symmetry property, which allows us to compare the size of the bias between two mirrored sequences. In total, 152 sequences occurred at least once with their exact mirrored sequence within the same group. Does the size of the bias depend on whether the overall trend of the sequence was upwards or downwards? Or does it depend on whether the last movement was UP or DOWN?

[Insert Table 4 about here.]

A small number of the sequences who occurred together with their mirrored image within the same group of subjects were repeated once or even twice within the same group. In such cases we calculated the average response of each subject. To see whether the direction or general trend of the sequence affects the bias, we took the difference of the SDV response to the upward trending sequence minus the SDV to the mirrored one. Here upward trending was defined by the number of upward movements. A sequence is defined upward trending if more than four outcomes were 'UP' or, in case exactly four outcomes were UP, if the last outcome was 'UP'.[16] Table 4 shows the results of random effects regressions on a constant of the resulting data, again controlling for group and subject effects. All results prove that the general direction of a sequence has no effect on reported biases. The same holds if one takes the difference of the SDV's not according to the general direction but according to the last outcome only.

Individual Characteristics

At the end of the experiment while the payments were distributed, subjects answered a set of questions. This endowed us with socio-demographic variables like age, mother tongue (English of foreign), sex and field of studies. None of these variables turned out to be correlated to the submitted valuations or biases.

[16]Excluding cases with exactly 4 'UP' movements from the analysis does not change results.

6. Conclusions

Based on previous experimental evidence, I constructed a prediction task experiment in which subjects are asked to predict the outcome of the next step of a sequence of eight i.i.d. random draws, represented by the repeated flipping of a fair coin. Contrary to existing studies the elicitation method has positive expected value at the normative prediction, does not rely on tournament incentives and instructions truthfully explain the underlying process as the outcomes of flips of a fair coin. A process which should easily be understood by subjects. The mechanism is incentive compatible even under some forms of risk preferences and symmetric around the normative prediction irrespective of risk attitudes.

The experiment tests the robustness of existing evidence of biases in the perception of independence and explores whether they are robust with respect to market feedback. As the main interest lies on the stability of individual biases rather than market biases, subjects don't interact with each other in a market but rather receive information about the median valuation in their group. This would be the market price if the implemented prediction task experiment was designed as a market where traders only submit their indifference price and the auctioneer set the market clearing price accordingly.

Observed experimental behavior in the prediction task experiments confirm previously made observations concerning biases in the perception of *randomness*. Subject's belief in a history dependent self correction of an i.i.d. process. This is especially reflected in an increasing belief in a reversal if several of the previous outcomes were identical. Similarly, an outcome is believed to be become more likely, the less often it occurred in the last number of realizations. The frequency of reversals has a significant negative effect, implying that if the sequence had many alternations, a further alternation is expected to be more likely.

Testing whether these biases are stable if subjects receive market feedback in the form of the median valuation, we find that biases are reduced, though only insignificantly. This is the case despite the fact that the majority of the received feedback was at the normative prediction and despite a rather large number of 98 paid repetitions.

With respect to the model by Barberis et al. (1998) our results have important implications. In their model they essentially assume that traders think that there are two different possible regimes which only differ in the switching probabilities of equidistant up and down movements. Agents adjust their belief about in which regime they are according to Bayes rule what implies that the more often the same outcome occurred the more an agent's belief in a continuation regime is strengthened. Our results clearly reject that as with increasing streak length subjects increasingly predict a reversal. However, the small but significant effect on high numbers of reversals confirms the model. Note however, that their model is not directly applicable as they assume that subjects do not have perfect information about the underlying process, as was the case in our experiment.

This limited applicability applies even more to the model by Rabin (2002). Here subjects infer the overall frequency of the outcomes from the observed data, suffering from the erroneous belief that within a moving band of outcomes draws are made from an urn with a small limited number of balls without replacement. The authors show that this may explain both, a Gambler's and a hot hand fallacy. As our data does not reflect a hot hand effect, this suggests that also their model is not confirmed by our data.

An important question which this study can not answer is whether biases survive in a proper market interaction. This paper only looked at the stability of individual biases. Future research must show whether these are traded away in a market such that the markets themselves are devoid of such biases.

References

Asparouhova, E., M. Hertzel, and M. Lemmon (2005). Behavioral biases and investor behavior: Predicting the next step of a random walk (revisited and extended). Workingpaper.

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

Bar-Hillel, M. and W. A. Wagenaar (1991). The perception of randomness. *Advances in Applied Mathematics 12*, 428–454.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343.

Becker, G., M. DeGroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioural Science 9*, 226–232.

Bloomfield, R. and J. Hales (2002). Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs. *Journal of Financial Economics 65*(3), 397–414.

Bolle, F. (1988). Learning to make good predictions for the following value in a timeseries. In W. Albers, R. Selten, and R. Tietz (Eds.), *Modeling Bounded Rational Behavior in Experimental Games and Markets*, Volume 314 of *LectureNotes*, pp. 37–50. Springer.

Bremer, M. and K. Kato (1996). Trading volume for winners and losers on the tokyo stock exchange. *The Journal of Financial and Quantitative Analysis 31*(1), 127–142.

Budescu, D. V. (1987). A markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance 13*, 25–39. Human Perception and Performance.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: An experimental study. *The Journal of Political Economy 95*(1), 1–33.

Camerer, C. F. (1987). Do biases in probability judgment matter in markets? experimental evidence. *American Economic Review 77*(5), 981–997.

Camerer, C. F. (1989). Does the basketball market believe in the 'hot hand,'? *The American Economic Review 79*(5), 1257–1261.

Camerer, C. F. (1992). The rationality of prices and volume in experimental markets. *Organizational Behavior and Human Decision Processes 51*(2), 237–272.

Chevalier, J. and G. Ellison (1997). Risk-taking by mutual funds as a response to incentives. *Journal of Political Economy 105*(6), 1167–1200.

Chevalier, J. and G. Ellison (1999). Are some mutual fund managers better than others? cross-sectional patterns in behavior and performance. *Journal of Finance 54*(3), 875–899.

Clotfelter, C. T. and P. J. Cook (1993). The "gambler's fallacy" in lottery play. *Management Science 39*(12), 1521–1525.

Dwyer Jr., G. P., A. W. Williams, R. C. Battalio, and T. I. Mason (1993). Tests of rational expectations in a stark setting. *The Economic Journal 103*(418), 586–601.

Falk, R. (1981). The perception of randomness. *Proceedings of the 5th International Conference for the Psychology of Mathematical Education*, 222–229. Grenoble, France.

Gardner, E. S. J. and E. McKenzie (1985). Forecasting trends in time series. *Management Science 31*(10), 1237–1246.

Gilden, D. L. and S. G. Wilson (1995). On the nature of streaks in signal detection. *Cognitive Psychology 28*, 17–64.

Gilden, D. L. and S. G. Wilson (1996). Streaks in skilled performance. *Psychonomic Bulletin & Review 2*, 260–265.

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

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003*, Volume 63 of *GWDG Bericht*. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Guryan, J. and M. S. Kearney (2008). Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review 98*(1), 458–473.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*, 263–291.

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review 102*(3), 490–502.

Koessler, F., C. Noussair, and A. Ziegelmeyer (2006). Parimutuel betting under asymmetric information. mimeo.

Lopes, L. L. and G. C. Oden (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology 13*(3), 392–400.

Metzger, M. A. (1985). Biases in betting: An application of laboratory findings. *Psychological Reports 56*(3), 883–888.

Miller, G. A. (1956). The magical number 7 plus or minus two: Some limits on our capacity in processing information. *Psychological Review 63*, 81–97.

Neuringer, A. (1986). Can people behave "randomly"?: The role of feedback. *Journal of Experimental Psychology 115*(1), 62–75.

Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance 53*(5), 1775–1798.

Ottaviani, M. and P. N. Sørensen (forthcoming). Surprised by the parimutuel odds? *American Economic Review*. mimeo.

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

Rabin, M. and D. Vayanos (2007). The gambler's and hot-hand fallacies: Theory and applications. *CEPR Discussion Papers 6081.*

Rapoport, A. and D. V. Budescu (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General 121*(3), 352–363.

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

Robson, A. J. (2002). Evolution and human nature. *The Journal of Economic Perspectives 16*(2), 89–106.

Schmalensee, R. (1976). An experimental study of expectation formation. *Econometrica 44*(1), 17–41.

Shefrin, H. and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance 40*(3), 777–790.

Sundali, J. and R. Croson (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making 1*(1), 1–12.

Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty 8*(3), 309–317.

Tversky, A. and D. Kahneman (1971). Belief in the law of small numbers. *Psychological Bulletin 76*(2), 150–110.

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin 77*, 65–72.

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

Williams, L. V. (1999). Information efficiency in betting markets: a survey. *Bulletin of Economic Research 51*(1), 1–39.

A. Analysis of Mechanism

In the following, without loss of generality, the payoff of the lottery will be set to 1 and 0, respectively. Denote by p^e the subjective probability that the outcome of the lottery will be 1. The submitted valuation of the lottery is denoted by v and r is drawn from a uniform distribution with $r \in [0, 1]$. Lets assume that preferences over outcomes are complete, transitive, and monotone, and can be represented by a utility function $u : \mathbb{R} \to \mathbb{R}$ with $u'(x) > 0$. Then, we can denote the expected utility, given a subjective probability p^e and a submitted valuation v by:

$$
\begin{aligned}
E[U|p^e, v] &= \int_0^v [p^e u(1-r) + (1-p^e)u(-r)]\mathrm{d}r + && (1)\\
&\qquad \int_v^1 [p^e u(r-1) + (1-p^e)u(r)]\mathrm{d}r && (2)\\
&= U(1) + U(0) - p^e[U(1-v) + U(v-1)] - (1-p^e)[U(v) + U(-v)] && (3)
\end{aligned}
$$

where $U(x)$ is defined by $\frac{\delta U(x)}{\delta x} = u(x)$. Maximizing with respect to v yields first order condition

$$\frac{p^e}{1-p^e} = \frac{u(v) - u(-v)}{u(1-v) - u(v-1)} \tag{4}$$

It is easy to see that it must hold that $v(p^e = 0) = 0$, $v(p^e = 0.5) = 0.5$, and $v(p^e = 1) = 1$ (to see this, simply plug in the respective values of v). As the right hand side of equation (4) is strictly increasing in v for all $0 \leq v < 1/2$, $1/2 < v \leq 1$, and utility functions with $u'(x) > 0$, we can state that every subjective probability is associated with a unique valuation and that whenever $p \in [0, 1/2]$ ($p \in [1/2, 1]$), it holds that $v \in [0, 1/2]$ ($v \in [1/2, 1]$).

We can furthermore state Theorem 1 (see section 4). To see this, let's assume that v satisfies condition (4) for some p^e_v. Now, compare this to a valuation $w = 1 - v$ which is optimal for p^e_w. By simple substitution of w with $1 - v$ it follows that $p^e_w = 1 - p^e_v$, proving the theorem.

It is also easy to see that under risk neutrality, i.e. for any linear utility function, it will hold that the utility maximizing valuation satisfies $v^* = p^e$, implying that the mechanism is incentive compatible under expected value maximization.

In general, the mechanism will be incentive compatible, i.e. $v^* = p^e$ whenever the utility function satisfies the following symmetry condition for all $v \in [0,1]$.

$$\frac{u(v) - u(-v)}{2v} = \frac{u(1-v) - u(v-1)}{2(1-v)} \tag{5}$$

To see this, rearrange to:

$$\frac{u(v) - u(-v)}{u(1-v) - u(v-1)} = \frac{v}{1-v} \tag{6}$$

which in combination with (4) implies $v^* = p^e$.

Here condition (5) compares the slopes of two lines, each connecting two points of the utility function: The left hand side of equation (5) connects point $(-x, u(-x))$ with $(x, u(x))$ and the right hand side connects $((x-1), u(x-1))$ with $((1-x), u(1-x))$. Note that there exist risk averse or risk seeking utility functions with $u''(x) \neq 0$ which satisfy symmetry condition (5) what implies that even under risk preferences, the mechanism can be incentive compatible. This is for example the case for $u(x) = x - \frac{1}{2}x^2$.

But what will be the bias in v if condition (5) is not satisfied? Assume that in condition (5) the slope on the left hand side is larger than the other one. Then, by simple calculations, we have $v^* < p$. Thus, the direction of the bias, i.e. whether it is directed towards 1/2 or towards the edges, depends on the relation between the two sides of symmetry condition (5). Independently of whether preferences reflect risk aversion or risk proness, both directions are possible.

B. Instructions Baseline Treatment

Introduction

Welcome to this experiment on individual decision making. In this experiment you will be able to earn money depending on your own decisions and random events. It is therefore of utmost importance that you read these instructions carefully. If you have any questions please raise your hand and wait for one of us to answer your questions in private. Please switch off your mobile and do not try to communicate with other participants. For showing up in time you will receive a show up fee of £5.

The experiment consists of a total of 100 rounds. The first two rounds are trial rounds to familiarize you with the experiment and will not be paid. In each of the remaining 98 rounds you will earn or potentially loose money. At the end of the experiment you will receive in private the sum of all round incomes minus all losses.

During the experiment all monetary amounts are denoted in "ECU" (*Experimental*

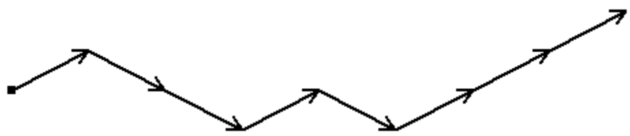
Currency Unit) which will be converted to British Pounds at the end of the experiment according to the following conversion rate

$$250 \text{ ECU} = 1 \text{ £}$$

To cover for potential losses you receive a flat payment of 500 ECU at the beginning of the experiment. If, however, at some point you accumulate more losses than 500 ECU you will only receive the show up fee.[17]

Procedure in each Round

In each of the following 100 rounds you will see a plot of a random process representing the outcomes of eight flips of one fair coin (the flips of the coin were recorded before the experiment). Here the outcomes of the flips of the coin are represented by UP or $DOWN$ movements of the graph. If the coin lands heads the graph moves UP if it lands tails it moves $DOWN$. Like with all fair coins, the chances of it landing heads or tails are equal. For example in the following graph the eight outcomes are (UP, $DOWN$, $DOWN$, UP, $DOWN$, UP, UP, UP).



What you don't see is the outcome of the ninth flip of the coin. This ninth flip of the coin is associated with a lottery which pays 100 ECU if the outcome is UP and nothing (0 ECU) if the outcome is $DOWN$. Your task is to submit a price for this lottery at which you are indifferent between buying a ticket for the lottery (paying the price and obtaining the outcome of the lottery), and selling a ticket (receiving the price and paying the outcome of the lottery).

After you have submitted your indifference price p, the computer generates a ticket price t for the lottery. This ticket price is a random number between 0 and 100 where all prices are equally likely. Knowing your indifference price p and the ticket price t the program then automatically buys or sells one lottery ticket for you. It buys one ticket if the ticket price is below your indifference price ($t < p$), and it sells one ticket if the ticket price is above your indifference price ($t > p$). More precisely

[17] Note that you would, nevertheless, have to stay until the end of the experiment!

- If $t < p$:
  If the ticket price t is smaller than your indifference price p (the lottery is cheaper than what you think it is worth), then the program **buys** one lottery ticket for you at price t and you will **receive** the outcome of the ninth flip of the coin (100ECU if outcome is UP and nothing if outcome is $DOWN$).
- If $t > p$:
  If the ticket price t is larger than your indifference price p (the lottery is more expensive than what you think it is worth), then the program **sells** one lottery ticket for you at price t and you will have to **pay** the outcome of the ninth flip of the coin (100ECU if outcome is UP and nothing if outcome is $DOWN$).
- If $t = p$:
  Finally, if the ticket price equals your indifference price nothing happens. You neither buy nor sell one ticket nor do you receive or pay the outcome of the lottery.

Round Payoff

Therefore, depending on whether the ticket price t is smaller or larger than your indifference price p and depending on the outcome of the ninth flip of the coin, your payoff is determined as follows:

- If $t < p$ (the ticket is cheaper than your indifference price)

  If the outcome is UP you receive $100 - t$

  If the outcome is $DOWN$ you **pay** t (in other words your payoff is $0 - t$)
- If $t > p$ (the ticket is more expensive than your indifference price)

  If the outcome is UP you **pay** $100 - t$ (in other words your payoff is $t - 100$)

  If the outcome is $DOWN$ you receive t

The reason why we implement this mechanism is that it implies that someone who wants to maximize his expected payoffs should always report what he thinks the lottery is worth. Why is that so? Remember that whenever you buy one lottery ticket the ticket price you pay is below your indifference price and whenever you sell one lottery ticket the ticket price you obtain is always higher than your indifference price. So you always pay less or obtain more than what you think the lottery is worth.

If, however, you would submit an indifference price p below what you think the lottery is worth it can happen that you sell the lottery for less than what you think it is worth.

Conversely, if you would submit an indifference price above what you think the lottery is worth, then it can happen that you buy the lottery ticket for a price higher than what you think it is worth.

At the end of each round you are informed about the random ticket price, the outcome of the ninth flip of the coin and your payoff.

Note: During the experiment you may have to wait until other participants have submitted their decision. This is purely for technical reasons. Neither do your decisions affect the payoff of others, nor do the decisions of others affect your payoff.

C. Instructions Feedback Treatment

The instructions for the feedback treatment were identical up to the last two paragraphs which read instead:

...

At the end of each round you are informed about the random ticket price, the outcome of the ninth flip of the coin and your payoff. You will also be informed about the median of the indifference prices submitted by a total of five participants including yourself. More specifically you are grouped with four other participants and all of you will see the same outcomes of the flips of the same coin in every round. At the end of every round all of you are informed about the median of the five indifference prices submitted by the participants in your group.

What is the median? Take for example the following five numbers: 1234, 567, 890, 987 and 654. The median of these five numbers is 890. In other words, if we sort these numbers from smallest to largest as follows: 567, 654, 890, 987, 1234 the third highest number is the median because two numbers are higher and two are smaller.

You will not see the median in the first two trial rounds!

Note: During the experiment you will have to wait until the other four participants have submitted their decisions. This is only so that the median can be computed and shown to you. Neither do your decisions affect the payoff of others, nor do the decisions of others affect your payoff.

List of Figures

Figures

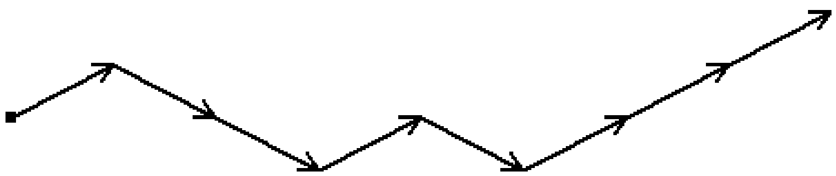


Figure 1: Representation of a Sequence in the Experiment

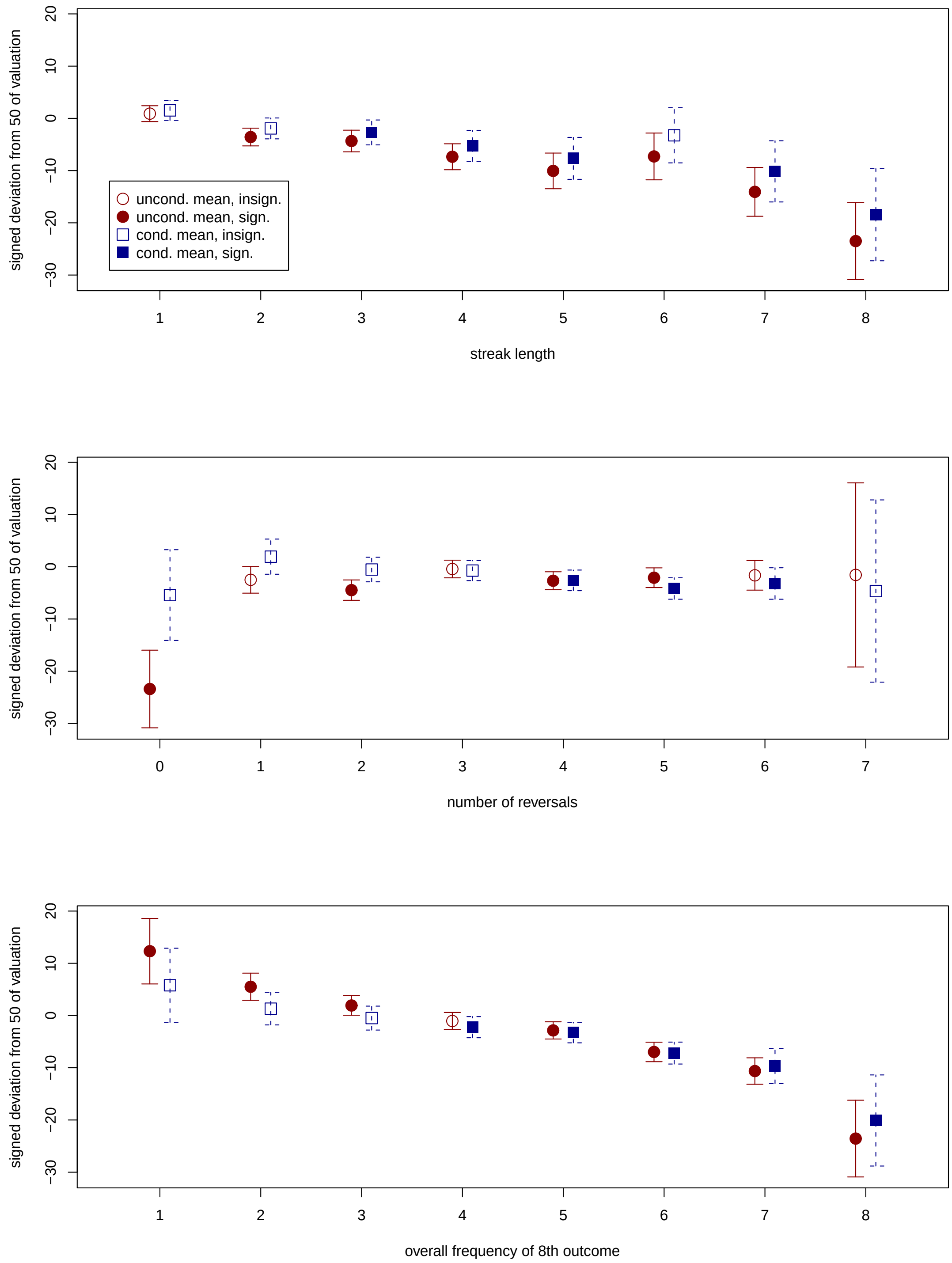


Figure 2: Mean Signed Deviation of Valuations

List of Tables

Tables

Table 1: Linear Mixed Effects Models of Signed Deviation by Treatment

	Info = 0		Info = 1		
	(1)	(2)	(3)	(4)	(5)
Intercept	-1.764	-1.784	-5.064***	-3.721***	-2.587**
	(1.219)	(.9869)	(1.714)	(1.178)	(1.020)
*dm*Streak	-2.748***	-2.739***	-1.128**	-1.131***	-1.154***
	(.6600)	(.6614)	(.4479)	(.4481)	(.4463)
*dm*Reversals	-1.591***	-1.597***	-.8480*	-0.853*	-0.839*
	(.5435)	(.5439)	(.4242)	(.4241)	(.4258)
*dm*Freq8th	-1.846*	-1.867*	-2.559***	-2.570***	-2.544***
	(.8450)	(.8460)	(.9877)	(.9890)	(.9877)
Period	-0.000	–	0.021	0.023	–
	(.0169)		(.0119)	(.0119)	
Median$_{t-1}$	–	–	0.0283	–	–
			(.0262)		
σ(resid.)	27.794	27.828	21.657	21.658	21.669
$\log L$	-21479	-21480	-19749	-19750	-19751
Wald χ^2(DF)	29.61(4)***	29.60(3)***	21.69(5)***	20.49(4)***	16.89(3)***

NOTE: Estimates (standard errors) of mixed effects models with random effects per group and per subject nested in group. Subject effects include random effect for every coefficient in regression. Model 1 and 2: 10 groups, 46 subjects with 98 observations each. Models 3 to 5: 9 groups, 45 subjects with 97 (98) obs. each in model 3 & 4 (5). Hausman tests confirmed random effects at $p = 5\%$. Significance levels: *** 1%, *** 2.5%, * 5%.

Table 2: Linear Mixed Effects Models of Signed Deviation by Treatment II

	Info = 0		Info = 1	
	(1)	(2)	(3)	(4)
Intercept	-1.384	-0.417	-4.513***	-3.599**
	(1.615)	(1.206)	(1.865)	(1.406)
*dm*Streak> 0	-2.048***	-1.359*	-0.959	-1.009*
	(.5973)	(.6855)	(.5438)	(.4731)
*dm*Streak< 0	-4.756***	-5.384***	-1.998**	-2.0700*
	(1.063)	(1.321)	(.8627)	(.9994)
*dm*Reversals> 0	-2.079***	-2.977***	-0.551	–
	(.7523)	(.8403)	(.6273)	
*dm*Reversals< 0	-1.076	–	-1.463*	-1.694*
	(.8663)		(.7196)	(.7920)
*dm*Freq8th> 0	-2.809***	-3.359***	-3.046***	-2.905***
	(.8174)	(1.172)	(1.130)	(1.101)
*dm*Freq8th< 0	-0.906	–	-2.014	-2.308
	(.7732)		(1.113)	(1.291)
Period	-0.007	–	0.022	–
	(.0153)		(.0119)	
Median_{t-1}	–	–	0.028	–
			(.0262)	
σ(resid.)	29.04	28.05	21.65	21.64
$\log L$	-21586	-21516	-19748	-19981
Wald χ^2(DF)	149.23(7)***	45.54(4)***	24.37(8)***	27.52(5)***

NOTE: Estimates (standard errors) of linear mixed effects models with random effects per group and per subject nested in group. Subject effects include random effect for every coefficient in regression. Model 1 and 2: 10 groups, 46 subjects with 98 observations each. Models 3 and 4: 9 groups, 45 subjects with 97 (98) obs. each in model 3 (4). Hausman tests confirmed random effects at $p = 5\%$. Significance levels: *** 1%, *** 2.5%, * 5%.

Table 3: Linear Mixed Effects Model of Signed Deviation - Both Treatments

model	*best*	*full*		*full (cnt.)*
Intercept	-2.413*	-1.688	D_I	-1.460
	(1.193)	(1.644)		(2.337)
dmStreak>0	-1.459***	-1.892***	$D_I\,dm$Streak>0	0.878
	(.4482)	(.6213)		(.8755)
dmStreak<0	-3.531***	-4.858***	$D_I\,dm$Streak<0	2.673
	(.8887)	(1.236)		(1.754)
dmReversals>0	-1.351***	-2.143***	$D_I\,dm$Reversals>0	1.555
	(.5079)	(.7220)		(1.009)
dmReversals<0	-1.330**	-1.195	$D_I\,dm$Reversals<0	-0.271
	(.5835)	(.8213)		(1.164)
dmFreq8th>0	-2.896***	-2.901***	$D_I\,dm$Freq8th>0	-0.183
	(.7764)	(.7773)		(1.555)
dmFreq8th<0	-1.628	-1.031	$D_I\,dm$Freq8th<0	-1.197
	(.8773)	(1.222)		(1.724)
Period	–	0.001	D_IPeriod	0.015
		(.0133)		(.0189)
σ(resid.)	24.88			24.88
$\log L$	-41612			-41608
AIC	83257			83268

NOTE: Estimates (standard error) of linear mixed effects models. Random effects per group and per subject nested in group. Subject effects include random effect for every base coefficient in the regression. 19 groups, 91 subjects with 98 observations each. Significance levels: *** 1%, *** 2.5%, * 5%.

Table 4: Delta SDV for Mirrored Sequences

		Trend	
ΔSDV	Info=0	Info=1	combined
Intercept	-0.9714	-1.992	-1.464
	(1.551)	(1.632)	(1.095)
		Last	
ΔSDV	Info=0	Info=1	combined
Intercept	0.794	-0.825	-0.022
	(1.551)	(1.111)	(0.951)
N	629	640	1269
Nsubj	46(10)	45(9)	91(19)

NOTE: Estimates (standard error) of linear random effects models with random effect on group and on subject nested in group. Data is difference of responses per subject between two mirrored sequences. Trend: difference is taken between the response to the upward trending sequence and the mirrored one. Last: difference is taken between the sequence where last outcome is ‘up’ and mirrored one. All results are insignificant with smallest $p = .222$.