What's Causing Overreaction?

An Experimental Investigation of Recency and the Hot Hand Effect*

Theo Offerman**

Joep Sonnemans**

Abstract
A substantial body of empirical literature provides evidence for overreaction in markets. Past losers outperform past winners in stock markets as well as in sports markets. Two hypotheses are consistent with this observation. First, the recency hypothesis states that traders overweigh recent information. Thus, they are too optimistic about winners and too pessimistic about losers. Second, the hot hand hypothesis states that traders try to discover trends in the past record of a firm or a team, and thereby overestimate the autocorrelation in the series. An experimental design allows us to distinguish between these hypotheses. The evidence is consistent with the hot hand hypothesis. Experience slightly reduces the observed phenomenon of overreaction.

JEL codes: D84, G12, C91.
Key-words: overreaction, recency, hot hand, scoring rule, experiments.

Current version: January 2000

** University of Amsterdam
CREED / Department of economics
Roetersstraat 11
1018 WB Amsterdam
Netherlands
e-mail: theoo@fee.uva.nl
e-mail: joeps@fee.uva.nl

* We would like to thank Uri Gneezy, Jan Potters, Arthur Schram, Randolph Sloof and Jörgen Wit for useful comments. The authors gratefully acknowledge financial support from NWO. Part of this work was completed while Theo Offerman worked at Tilburg University.

1. Introduction

In the economics profession the proposition that markets yield efficient outcomes has been considered a truism for a long time. For financial markets efficiency implies (among others) that prices resemble intrinsic values of stocks and that the development of the price of a stock over time is not distinguishable from a random walk. If the development of prices were predictable, arbitrageurs would detect the trends and make money. Their actions would immediately drive stock prices back to intrinsic values.

The view that financial markets are efficient has been seriously challenged in recent empirical work. De Bondt and Thaler (1985) report that prices in the New York Stock Exchange overreact in the period between 1926 and 1982. Stocks of firms that were doing (extremely) badly for the last 3 years are undervalued and stocks of firms that were doing (extremely) well are overvalued. Prior losers provide higher returns than prior winners in the following years, contradicting the random walk hypothesis. Subsequent empirical studies support this result (*e.g.*, De Bondt and Thaler, 1987; Chan, 1988; Fama and French, 1988; Chopra, Lakonishok and Ritter, 1992; see also the survey by Forbes, 1996). Often this phenomenon of overreaction is even more pronounced in stock markets outside the USA (*cf.* Poterba and Summers, 1988; Alonso and Rubio, 1990; Stock, 1990; da Costa, 1994). It may be tempting to think that it is only the misguided behavior of a small group of noise traders that is causing overreaction in stock markets. However, De Bondt and Thaler (1990), Bulkley and Harris (1997) and Amir and Ganzach (1998) present evidence that even the judgments of professional security analysts reflect systematic biases.

A rational (efficiency preserving) explanation for the phenomenon is provided by risk aversion. Losing firms are more likely the riskier firms. Therefore, rational traders would need an extra premium in order to buy their stocks. Although there is evidence that stocks of losing firms are riskier than stocks of winning firms, the majority of the aforementioned authors feels that the difference in risk is not large enough to explain the difference in expected returns.[1]

Instead, De Bondt and Thaler (1985, 1987) attribute overreaction to the psychological phenomenon of recency. When processing information, people tend to overweigh recent information compared with their prior belief. Thus, traders who are not sure of the intrinsic value of a stock will be too optimistic about its value when the firm is winning and too pessimistic when it is losing. Recency

[1] It has also been suggested that the effect of overreaction is confounded with the small firm effect. Small firms earn relatively high returns. However, De Bondt and Thaler (1989) report that the losers of the overreaction effect are not exactly the same small firms associated with the small firm effect. Chopra *et al.* (1992) find an economically important overreaction effect even after adjusting for risk and size.

may thus cause a temporary wedge between stock prices and intrinsic values.

A related literature on judgmental biases in sports markets is inspired by the article of Gilovich, Vallone and Tversky (1985). They show that both basketball players and fans believe that a player is more likely to hit a shot if his previous shot was a hit instead of a miss. However, they do not find a positive correlation between the actual outcomes of successive shots. Although Gilovich *et al.* use a different terminology, this bias is in principle similar to the phenomenon of overreaction in stock markets.

Camerer (1989) investigates whether a false belief in autocorrelation is also revealed in the betting market for basketball teams. Consistent with Gilovich *et al.*, he finds that winning teams are overvalued and that losing teams are undervalued in the NBA games between 1983 and 1986. However, the bias is too small to provide incentives for arbitration (due to transaction costs).[2] Overreaction may be more pronounced in other sports markets. Badarinathi and Kochman (1994) report results that betting against winning teams was broadly profitable for the NFL 1983-1992 football games. Their result is partly supported by the analysis of Tassoni (1996). He reports overreaction for the NFL football games in the periods 1956-1965 and 1976-1979, but not in the period 1980-1985.

In sports markets the phenomenon of overreaction is usually attributed to a mistaken belief in a 'hot hand'. Bettors believe that the performance of a winning (losing) team during a particular period is better (worse) than its overall record. They conclude too easily that there are trends in a team's past record by overestimating the autocorrelation in the results of a team's successive games. A similar judgmental bias is often found in the beliefs of gamblers in a casino. Knowing that the probability of red is equal to 0.5, they expect more alternations between red and black than would statistically be expected on the basis of pure chance (the gambler's fallacy is discussed by Tversky and Kahneman, 1982, and Terrell, 1994). Both gamblers falling prey to the gamblers' fallacy and bettors misled by the hot hand effect have the tendency to expect too many runs in a series given a certain amount of autocorrelation.

In stock markets traders form beliefs about the future value of the stocks. In sports markets bettors form beliefs about the future performance of teams. Stock market traders and sports markets bettors face a similar task. Furthermore, in stock markets stocks of winning firms are overvalued and stocks of losing firms are undervalued. In sports markets winning teams are overvalued and losing teams are undervalued. Thus in both types of markets the same phenomenon of overreaction is observed.

[2] Brown and Sauer (1993) also analyze the data of NBA games. Their results show that bettors believe in positive autocorrelation. However, they conclude that the data are not sufficiently informative to determine whether an actual correlation between team's successive performances does or does not exist, so it is not clear whether this belief is a bias or not.

Nevertheless, the phenomenon is explained differently in stock markets than in sports markets.

Note that the explanation for overreaction in stock markets (recency) can also be used to explain overreaction in sports markets: it could be that bettors in a sports market, uncertain of the strength of their team, overweigh recent information about the performance of the team. Recency implies that the prior belief does not receive enough weight in the updating process. In that case, bettors would be too optimistic about the value of winning teams and too pessimistic about the value of losing teams.

On the other hand, the explanation of overreaction in sports markets (hot hand) can also be used to explain overreaction in stock markets: it could be that traders try to discover trends in the past record of a firm. In doing so, they could overestimate the amount of autocorrelation. Assume for example that the time series of a stock is represented by a random walk. Believers in the hot hand effect would expect more alternations than they actually observe in a random walk. This would lead them to the false conclusion that they can detect whether a firm is in a good or a bad shape. Thus, they would put too much value on winning firms and too little value on losing firms.

These two conceptually different explanations for overreaction can obviously not be separated using either real world data of sports markets or of stock markets, since they both yield the same bias. It is possible to distinguish between these hypotheses in an experimental setting, however. The main goal of this paper is to determine experimentally whether recency or hot hand is the better explanation of overreaction.

Consider the following experimental setup. A coin is selected randomly from an urn containing an equal number of false and fair coins (the prior probability of a false coin = 0.5). A fair coin has no memory: each toss of the coin will be head (tail) with probability 0.5 (0.5). A false coin has the property that the previous outcome is repeated with probability 0.7. If the previous outcome was head, the new outcome will be head with probability 0.7 and it will be tail with probability 0.3. Thus, the outcome of the toss of a false coin depends only on the outcome of the previous toss. The outcome of the first toss with a false coin is head (tail) with probability 0.5 (0.5).

The decision maker is not told whether the randomly selected coin is fair or false. The coin is tossed twenty times yielding a series of heads and tails. The decision maker observes the series and predicts the probability that the series was generated by a false coin. The payoff of an incentive compatible mechanism (the quadratic scoring rule, *cf.* Murphy and Winkler, 1970) encourages the decision maker to take the task seriously.

How would a Bayesian observer handle this problem? In the following, let B denote the Bayesian posterior probability that the coin is false, let *P[false coin] (P[fair coin])* denote the prior probability that the coin is false (fair) and let y denote the number of alternations in the series. Then,

$$B = \frac{P\ [false\ coin] * P\ [data \mid false\ coin]}{P\ [false\ coin] * P\ [data \mid false\ coin] + P\ [fair\ coin] * P\ [data \mid fair\ coin]}$$

$$= \frac{0.5 * 0.7^{20-y-1} * 0.3^{y}}{0.5 * 0.7^{20-y-1} * 0.3^{y} + 0.5^{20}} * 100\% = \frac{100\%}{1 + (\frac{5}{7})^{19} * (\frac{7}{3})^{y}} \ . \qquad (1)$$

Note that a Bayesian observer only needs to know the number of alternations in the series. This number directly determines the Bayesian posterior probability that a coin is false.

The hot hand effect would induce a decision maker to overestimate the autocorrelation in the series. A series actually generated by a fair coin with 0 autocorrelation will then be perceived as a series of a false coin with positive autocorrelation. A series actually generated by a false coin with positive autocorrelation will then be perceived as a series of a false coin with even higher autocorrelation. In both cases, the decision maker reports a higher than Bayesian posterior probability that the coin is false. Recency occurs if a decision maker overweighs recent information (the series of coin tosses) and underweighs the prior information. The effect of recency depends on the number of alternations in the series of heads and tails generated by the coin. If this number is such that a Bayesian observer would report a higher probability than 50%, then neglect of the prior distribution would induce a decision maker to overestimate the probability that the coin is false. On the other hand, if the number of alternations leads a Bayesian observer to predict a probability smaller than 50%, then neglect of the prior distribution would induce a decision maker to underestimate the probability that the coin is false.[3] Thus, series that seem to be generated by a fair coin can be used to separate recency from hot hand.

Two other systematic biases in decision makers' judgments are possible. First, the decision maker could underestimate the autocorrelation in the series. Such a decision maker would fall prey to the so-called cold hand effect, and would make errors opposite to the errors predicted by the hot hand effect. Second, a decision maker could underweigh new evidence compared with the prior belief. Such a decision maker would be affected by conservatism, and make errors in the opposite direction of the errors implied by recency.[4] The design allows to discriminate between these four hypotheses. Together

[3] Grether (1980; 1992) nicely makes the point that subjects using the representativeness heuristic put too much weight on new information. Grether obtains evidence consistent with this heuristic in a series of experiments where he varies the induced prior belief.

[4] Edwards (1968) discusses some early experimental studies suggesting that biases in subjects' beliefs result from conservative updating. Adelman, Bresnick, Black, Marvin and Sak (1996) find that experienced Patriot officers overweigh prior information when identifying aircrafts as friendly, hostile or unknown.

they (almost) exhaust the possible systematic errors that could be made by a decision maker. Table 1 summarizes the predictions.

TABLE 1
Predictions

hypothesis	B<50%	B>50%
hot hand	R>B	R>B
recency	R<B	R>B
cold hand	R<B	R<B
conservatism	B<R<50%	50%<R<B

Notes: 'B' denotes the Bayesian posterior probability that the coin is false; 'R' denotes the decision maker's reported posterior probability that the coin is false.

This paper describes the results of two experiments. Experiment 1 is the basic experiment designed to address the question which of the hypotheses explains subjects' beliefs best. Our results show that the hot hand hypothesis gives a better account of the data than the recency hypothesis. Subsequently, in experiment 2 we investigate the robustness of the phenomenon of overreaction. There, subjects receive training before the start of the experiment.

The remainder of this paper is organized as follows: section 2 describes the experimental design in more detail. Section 3.1 provides the results for experiment 1. Section 3.2 presents the results for experiment 2. Section 4 contains a concluding discussion.

2. Experimental design

Both the instructions and the decision phase of the experiment are computerized.[5] The experimental situation is explained as follows:

> 'There is an urn containing fair and false coins. One coin will be drawn from this urn and this coin will be tossed repeatedly. You will observe the outcomes and we will ask you to estimate the probability that the coin is false. The fair and the false coins differ in the following sense: if the coin is fair, the probability of 'head' (H) and the probability of 'tail' (T) is 50% for each toss. A fair coin does not have a 'memory'. A false

[5]The program is developed in Turbo Pascal using the RatImage library (see Abbink and Sadrieh, 1995 for documentation of this library). The program is available from the authors.

coin does have a memory. The first toss with a false coin will yield outcome H with a probability of 50% and outcome T with a probability of 50%. After the first toss, the outcome of the previous toss will be repeated with a probability of 70%. So, if the coin is false and the previous toss yielded outcome T, the next toss will yield outcome T with a probability of 70% and outcome H with a probability of 30%. Likewise, if the previous toss yielded outcome H, the next toss will yield outcome H with a probability of 70% and outcome T with a probability of 30%.

The urn contains 50 false and 50 fair coins. The coin will be drawn randomly from this urn. Thus, the probability that an arbitrary coin is false is 50%. With the help of the computer this coin will be tossed 20 times. You will see the outcomes on your computer screen. Then we will ask you to estimate the probability that this coin is false.

This procedure will be repeated 20 times, so for 20 series of outcomes you have to estimate the probability that the series is tossed with a false coin. Each time the coin is drawn from an urn with 50 false and 50 fair coins. You will not see the urn. The computer will take care of the drawing of the coin and the tossing of that coin.'

Then it is explained how subjects make money. For each estimate they receive a payoff determined by a quadratic scoring rule. Let R denote the reported probability of a false coin in percentages, then the payoff is $10000\text{-}R^2$ points if the coin is fair and $200*R\text{-}R^2$ points if the coin is false. At the end of the experiment the points are exchanged for money (8000 points = 1 Dutch guilder).

For expected value maximizers the quadratic scoring rule is an incentive compatible mechanism to measure expectations. It has been used in McKelvey and Page (1990) to elicit subjects' beliefs in an information aggregation experiment. It has also been used in Offerman, Sonnemans and Schram (1996) and Sonnemans, Schram and Offerman (1998) to elicit subjects' beliefs in public good games. In the present experiment, subjects do not know the formula of the scoring rule, but receive a payoff table based on the formula on paper. The table displays the payoff for each (integer) estimate between 0% and 100% when the coin is fair and when the coin is false. The instructions explain that it is in the best interest of subjects to report their true beliefs. Subjects answer some questions to check their understanding.

All subjects observe the same 20 series generated by 20 coins. The whole series of 20 outcomes is displayed at the top of the screen with a delay of 0.5 seconds between successive outcomes. At the bottom of the screen a window appears asking the subject to report the probability that the coin is false. The subject confirms her or his percentage and has to wait a few seconds, before (s)he receives the outcomes generated by the subsequent coin. Only after the last (20th) coin the true state of each coin and the earnings are communicated to the subjects. At the end of the experiment subjects fill in a questionnaire before they are paid privately. The coins and the series generated by each coin have been produced with the help of a random number generator: 11 of the 20 coins are fair coins. Details of the series of the coins are presented in the appendix.[6]

Experiment 2 is designed to investigate whether the biases from Bayesian updating are robust to the possibility of learning. In experiment 1 there is no opportunity for learning. Subjects estimate posterior probabilities without any experience with series produced by false and fair coins. Experiment 2 allows subjects to observe some series produced by fair and false coins before they provide their estimates for the series of the 20 coins. The practice series are presented in exactly the same way as the series for which subjects estimate probabilities. The difference is that after a series is presented in the training phase, subjects do not estimate a probability. Instead, it is revealed whether the series is produced by a fair or a false coin. Then a subject chooses whether (s)he wants to see the series produced by an additional coin, up to a maximum of 100 coins. To encourage learning we do not impose a cost on observing series of coins. If the subject indicates that (s)he has observed enough series, the real experiment starts. The second part of the experiment is exactly the same as experiment 1.

Subjects

In experiment 1 a total of 39 subjects participated in two sessions. 25 of the 39 undergraduates are students of economics and 14 are students of other departments; 9 are female and 30 are male. An average of 19.50 guilders was earned by subjects in about 45 minutes.[7] Only one subject stated in the questionnaire that he did not report his true beliefs in the experiment: for the first 10 coins he reports 50% each time and for the second 10 coins he reports alternately 0% and 100%. The data of this subject are excluded from the analyses. This decision does not have an (important) impact on the results. In experiment 2 22 subjects participated: 10 subjects major in economics and 8 in other fields (4 did not report their field); 8 subjects are female and 11 are male. The gender of the other 3 subjects is unknown to us. An average of 20.20 guilders was earned by subjects in experiment 2.

[6] To investigate whether our results are affected by the order in which the series of outcomes generated by the coins are presented, we use a different order of the coins in the two sessions of experiment 1. In the second session coins 1 and 11, 2 and 12, etc, of the first session are swapped. The difference in the order of the coins between the two sessions does not affect subjects' judgments. We could not find any (systematic) difference in the judgments reported in the two sessions. Therefore, we abstract from the order of the coins in the remainder of this paper: the data of the two sessions are simply combined.

[7] One guilder can be exchanged for approximately 0.5 US$.

3. Results

3.1 Experiment 1: which hypothesis explains overreaction best?

Experiment 1 is designed to discriminate between rival explanations for the phenomenon of overreaction. This section focuses on the results of experiment 1.

Subjects tend to estimate the probability that a coin is false higher than a Bayesian observer would (on average 59.6% versus 45.9%). To see whether the difference is significant, we count for each subject the number of times that (s)he reports a higher than Bayesian probability that the coin is false. A two-tailed sign test clearly rejects the hypothesis that this number is equal to 10 (half of the total number of coins), in favor of the hypothesis that the former is higher (n=38; p=0.00). Figure 1 displays the biases per subject. Only one subject provides more underestimations than overestimations of the probability that the coin is false. One subject is unbiased on average. All other subjects more often overestimate than underestimate the probability that a coin is false.

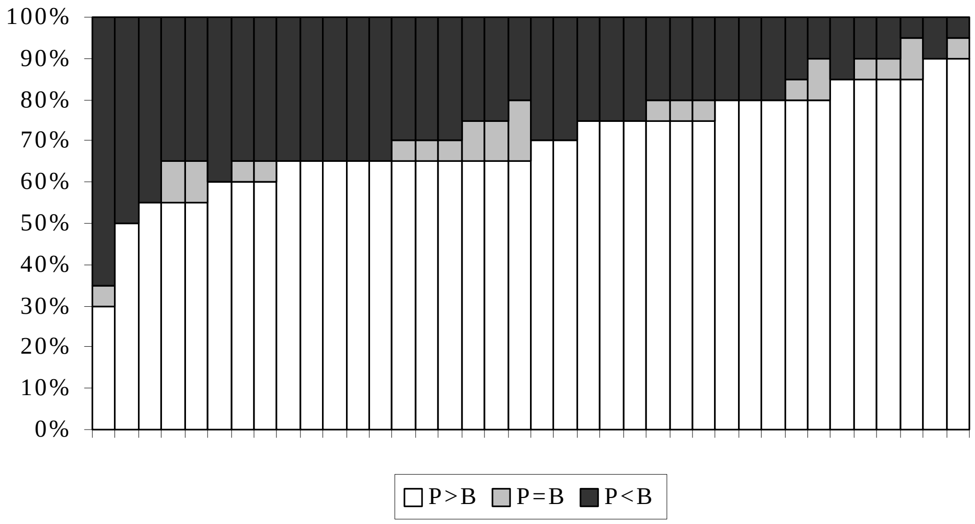

FIGURE 1: biases per subject in experiment 1. Each of the 38 subjects is represented by a bar. The upper (lower) part of the bar displays the percentage of coins for which a subject underestimated (overestimated) the probability that it was false; the middle part represents the percentage of accurate estimates.

Although judgments are biased, there is a clear positive correlation between reported and Bayesian probabilities (the Spearman rank correlation coefficient is equal to 0.62, p=0.00). Figure 2 plots the mean reported probabilities as function of the Bayesian probabilities. The figure shows that the difference between the reported and Bayesian probabilities increases when the Bayesian probability of a false coin decreases.

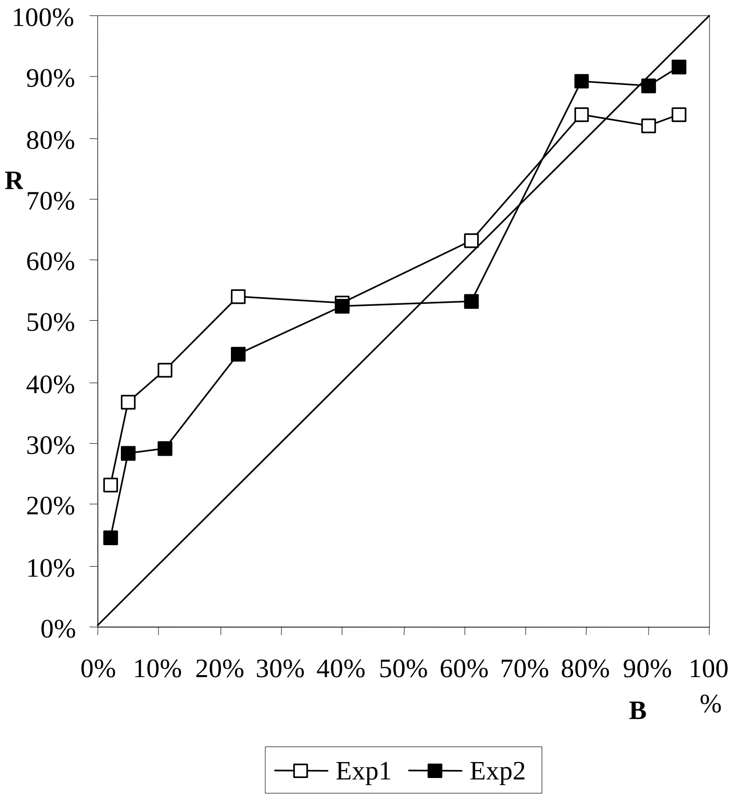

FIGURE 2: mean reported probabilities (R: vertical axis) in experiments 1 and 2 as function of the Bayesian probabilities (B: horizontal axis). The straight line R=B is added as a benchmark.

Despite the clear biases in their beliefs subjects do not lose that much money. Subjects earn 87.9% of what a Bayesian observer would earn in the experiment. The combination of considerable biases and a high efficiency level can be attributed to a well known characteristic of the scoring rule: the function is quite flat in the neighborhood of the optimum (*cf.* Davis and Holt, 1993, p.465-467).

A pair-wise comparison of the four hypotheses is made to assess which hypothesis takes best account of the biases in subjects' judgments. The results are summarized in table 2. For some comparisons we focus on a subset of the results. In those cases the hypotheses predict a similar bias for some coins. These coins are excluded from such a comparison. For example, in the comparison between hot hand and recency, we focus on coins that seem fair (B<50%), because both hypotheses predict that subjects will overestimate the probability of a false coin when the coin seems false (B>50%). For each subject it is counted how often (s)he judges in accordance with each hypothesis. Thus, for each comparison each subject yields one pair-wise data-point, and a Wilcoxon rank test is used to test the null that both hypotheses explain the biases equally well. The hot hand hypothesis beats all other hypotheses.[8]

TABLE 2
Pair-wise comparison of the four hypotheses

H1 versus H2	Which coins?	consistent first	consistent second	Wilcoxon Z-statistic
hot hand versus recency	B<50%	9.3	1.3	-5.27^{**}
hot hand versus cold hand	0%<B<100%	14.0	5.3	-5.10^{**}
hot hand versus conservatism	B>50%	4.7	3.1	-2.11^{*}
recency versus cold hand	B>50%	4.7	4.0	-0.91
recency versus conservatism	0%<B<100%	6.0	8.6	-2.25^{*}
cold hand versus conservatism	B<50%	1.3	5.5	-4.61^{**}

Notes: The second column indicates which coins are used to test the hypotheses in the first column; the third (fourth) column indicates how often a subject made a judgment in accordance with the first (second) hypothesis of the row; the last column shows the Wilcoxon rank test statistic (n=38 for all tests; * indicates significance at the 5% level and ** at the 1% level).

The biases are best explained by the hot hand hypothesis.[9] However, close examination of figure 2 reveals that the evidence is more clearcut for coins that seem fair than for coins that seem false. This might be explained by the presence of white noise: it is more likely that the positive bias caused by a hot hand effect is offset by a negative bias caused by white noise for coins with a high Bayesian probability of being false. To formalize the idea that the data are generated by the combination of a systematic error component provided by either of the four hypotheses and a random component provided by white noise, we estimate five models using maximum likelihood techniques. The base model maintains the hypothesis that subjects are Bayesian and that all errors are white noise:

$$R = B + e \; ; \qquad (2)$$

[8] We do not find any differences according to gender or major.

[9] A well known characteristic of the scoring rule is that risk averse and risk loving subjects should report somewhat different beliefs than their true beliefs. In Sonnemans and Offerman (1998) we investigate whether the results are affected by this characteristic of the scoring rule. After reporting 20 probabilities subjects make choices between gambles constructed on the basis of the scoring rule. The procedure employed allows us to map reported probabilities into 'true subjective' probabilities. On average the reported probabilities are indistinguishable from true subjective probabilities. We conclude that the scoring rule does not (systematically) bias reported beliefs.

Recall that R represents the reported and B the Bayesian posterior probability that a coin is false. For each model proposed it is assumed that all random error terms ε are independently drawn from the same truncated $N(0,\sigma^2)$ distribution. The distribution of the error terms is truncated, because subjects never report probabilities smaller than 0% or greater than 100% (*cf.* Mood, Graybill and Boes, 1985, p.124).

First, the hot hand model proposes:

$$R = B + \alpha * (100 - B) + \varepsilon \ . \tag{3}$$

If α is estimated to be equal to a value between 0 and 1, the equation implies that subjects overestimate the probability of a false coin for all coins. This would only be expected on the basis of the hot hand hypothesis.

Second, the recency model proposes:

$$R = \begin{cases} B * (1 - \beta) + \varepsilon & \text{if} \quad B < 50\% \\ B + \beta * (100 - B) + \varepsilon & \text{if} \quad B > 50\% \end{cases} \tag{4}$$

With a β between 0 and 1, the above equation implies that subjects overestimate the probability of a false coin for coins that seem false, but underestimate this probability for coins that seem fair as proposed by the recency hypothesis.

Third, the cold hand model assumes:

$$R = B * (1 - \gamma) + \varepsilon \ . \tag{5}$$

If γ lies between 0 and 1, this equation implies that subjects always underestimate the probability of a false coin. This would only be consistent with the cold hand hypothesis.

And finally, the conservatism model predicts:

$$R = B + \delta * (50 - B) + \varepsilon \ . \tag{6}$$

With a δ between 0 and 1, this equation implies that subjects make errors biasing their predictions towards 50% as predicted by conservatism. The base model is nested in each of the four models, by setting the parameter of that model (α, β, γ or δ) equal to 0. We also add the results for the random model as a benchmark. According to the random model each of the 101 possible percentages is selected with equal probability. The random model is nested in the base model, by letting $\sigma \rightarrow \infty$.

Table 3 summarizes the results of the estimation procedure. The base model explains the data significantly better than the random model. The hot hand model explains the data better than the base model. Besides a significantly better likelihood the hot hand model also provides a considerably better (smaller) estimate of the white noise component ó. The other general models do not explain the data significantly better than the base model. It is concluded that hot hand plus white noise provides the best explanation of the data.

TABLE 3
Maximum likelihood estimates

Model	parameter	ó	-logL
random	XX	XX	3507.5
base	XX	37.0	3394.0**
hot hand	á=0.30	26.8	3322.3**
recency	â=0.07	37.6	3393.7
cold hand	ã=0.00	37.0	3394.0
conservatism	ä=0.00	37.0	3394.0

Notes: The loglikelihood is computed on the basis of 760 choices. ** indicates significance of the likelihood ratio test at the 1% (comparison of the likelihood of the model with the nested model).

The foregoing analysis has been inspired by hypotheses suggested by the literature. However, one may wonder whether other factors or biases also play a role in determining a subject's reported probability of a false coin. We consider two possibilities: the posterior subjective probability of a false coin may be affected by the length of the maximal run in the series generated by the coin; the posterior subjective probability may be affected by a bias in the relative frequency of heads and tails in the series.

The number of alternations in the series is a sufficient statistic for a Bayesian observer (*cf.* equation 1). Thus, a Bayesian observer should allocate equal probabilities to coins with equal numbers of alternations. On the other hand, for non-statisticians a series may provide stronger evidence for a false coin if the length of the maximal run is longer. This seems to be the case. Twelve times a comparison can be made between two coins that generated an equal number of alternations but an unequal length of the maximal run. For 10 (2) of the 12 comparisons, subjects estimate the probability higher (lower) if the length of the maximal run is longer.

The other factor considered is the bias in relative frequency. Given a number of alternations, a series may provide a subject with stronger evidence for a false coin if the overall relative frequency of tails is further removed from 0.5. This factor does not seem to play a role. Of the 9 times that a comparison can be made between two coins with equal number of alternations but unequal relative frequency, only 5 predictions are in line with the direction predicted by this factor.

3.2 Experiment 2: the effect of training

In experiment 2 we investigate whether biases from Bayesian predictions disappear when subjects receive training before the start of the experiment. With training subjects still overestimate the probability that a series has been produced by a false coin (average estimate is 55.8% versus a 45.9% probability for the Bayesian observer). Again we count for each subject the number of times that (s)he reports a higher than Bayesian probability that the coin is false. A large majority of 20 of the 22 subjects reports a higher than Bayesian probability for more than half of the coins. A two-tailed sign test reveals that this number is significantly greater than 10 (n=22; p=0.00).

Nevertheless, prior experience weakens the bias against the Bayesian framework slightly (in experiment 1 the average probability is estimated to be 59.6%). Although the difference is small, the average probability reported by subjects in experiment 2 has a significantly lower rank than the average probability reported by subjects in experiment 1 (Mann-Whitney rank test: m=22, n=38; Z=-1.94; p=0.05). Subjects' earnings are also somewhat higher in experiment 2 than in experiment 1.

Figure 2 displays the average estimated probability of a false coin as a function of the Bayesian probability for experiments 1 and 2. The results of the two experiments are qualitatively the same. The main difference is that experience seems to help them evaluate fair coins somewhat better.

On average subjects observe 8.8 series generated by practice coins (minimum 2; maximum 25). There is only a weak correlation between number of coins practiced and average reported probability of false coin (-0.16, not significant). There is no correlation between number of coins practiced and earnings. Perhaps those who need practicing recognize that they need it, such that a potential positive learning effect is largely offset by a negative effect of worse skills.

6. Concluding discussion

Two psychological hypotheses potentially account for overreaction in stock and sports markets. Trading data of real markets cannot be used to separate these hypotheses. The simple experimental design in this paper provides an opportunity to differentiate between the two explanations. The data generated with this design are clearly better explained by the hypothesis stating that biases consist of a systematic part caused by hot hand and a random part caused by white noise. Recency cannot explain our data. In fact, the results reported in table 2 suggest that if a bias comes up in the weighing of new information, it is that people tend to put insufficient value to new information.

The psychological literature does not propose that people always underweigh or always overweigh new information. Hogarth and Einhorn (1992) classify a multitude of studies on order effects in belief learning. They propose that the interaction between the response mode (either a judgment is made after each piece of information, or a judgment is made after a series of information pieces), the complexity of the task (simple or complex) and the number of pieces of information (short or long series) determines whether recency or conservatism will be observed. Recency may not provide such a stable basis to predict biases in markets.

Both the recency explanation of overreaction in stock markets and the hot hand explanation of overreaction in predicting sports events were originally attributed to the more general phenomenon of representativeness (by De Bondt and Thaler, 1985 and by Gilovich *et al.*, 1986, respectively). According to Tversky and Kahneman (1982, p. 24), "people view a sample randomly drawn from a population as highly representative, that is, similar to the population in all essential characteristics. Consequently, they expect any two samples drawn from a particular population to be more similar to one another and to the population than sampling theory predicts, at least for small samples." The representativeness heuristic triggers different judgmental biases. For example, besides the recency effect (or insensitivity to prior probabilities) and the hot hand effect, representativeness also predicts judgments to be insensitive to sample size.

This paper suggests that the original concept of representativeness is imprecise, in the sense that it allows for countervailing forces. In the present design, the hot hand effect appears to be a stronger force than recency. Furthermore, in this study subjects seem to believe the evidence for the falseness of a coin is stronger if the length of the maximal run of its series is longer. These findings may help to bring more precision to the idea of representativeness.

The goal of this paper is to find the cause of overreaction in markets. In doing so, it turns out that individual judgmental biases in this design are considerable: on average subjects estimate the posterior probabilities of a false coin about 14% above Bayesian probabilities. This bias is only slightly reduced if subjects are allowed to obtain free experience with series generated by false and fair coins.

Of course, our experiment abstracts from some elements that may affect overreaction in markets. For example, biased traders may learn from unbiased traders via the signals provided by market prices. Or vigorous trading by unbiased investors might to some extent neutralize the effect of trading of biased traders on the aggregate price level. These are interesting topics for future experimental work, especially since other studies suggest that markets in some circumstances alleviate the effects of some judgmental biases (*cf.* Camerer, 1987; Camerer, Loewenstein and Weber, 1989; Anderson and Sunder, 1995; Ganguly, Kagel and Moser, 1998).

Knowledge about the cause of overreaction is of scientific interest because it improves the understanding of what we observe in markets. But it is also of practical relevance. Consider the case that a trader tries to form a belief about the value of a stock that was previously unknown to him or her. This paper suggests that (s)he should not be too concerned that (s)he overweighs recently revealed information about the stock. It suggests that (s)he should not conclude too easily that the stock is in a good or a bad shape on the basis of trends in the past record of the stock. More generally, the training of traders should focus more on the pitfall of perceiving trends too easily than on the mistake of overweighing recent information.

References

Abbink, K., and A. Sadrieh, 1995: 'Ratimage, Research Assistance Toolbox for Computer-Aided Human Behavior Experiments', *Discussion paper B-325*, University of Bonn.

Adelman, L., T. Bresnick, P.K. Black, F.F. Marvin, and S.G. Sak, 1996: 'Research with Patriot Air Defense Officers: Examining Information Order Effects', *Human Factors* 38, 250-261.

Alonso, A., and G. Rubio, 1990: 'Overreaction in the Spanish Equity Market', *Journal of Banking and Finance* 14, 469-481.

Anderson, M.J., and S. Sunder, 1995: 'Professional Traders as Intuitive Bayesians', *Organizational Behavior and Human Decision Processes* 64, 185-202.

Badarinathi, R., and L. Kochman, 1994: 'Does the Football market Believe in the "Hot Hand"?', (Anthology), *Atlantic Economic Journal* 22, 76.

Brown, W.O., and R.D. Sauer, 1993: 'Does the Basketball Market Believe in the Hot Hand?', *American Economic Review* 83, 1377-1386.

Bulkley, G., and R.D.F. Harris, 1997: 'Irrational Analysts' Expectations as a Cause of Excess Volatility in Stock Prices', *Economic Journal* 107, 359-371.

Camerer, C.F., 1987: 'Do Biases in Probability Judgments Matter in Markets? Experimental Evidence', *American Economic Review* 77, 981-997.

Camerer, C.F., 1989: 'Does the Basketball Market Believe in the 'Hot Hand'?', *American Economic Review* 79, 1257-1261.

Camerer, C., G. Loewenstein, and M. Weber, 1989: 'The Curse of Knowledge in Economic Settings: An Experimental Analysis', *Journal of Political Economy* 97, 1232-1254.

Chan, K.C., 1988: 'On the Contrarian Investment Strategy', *Journal of Business* 61, 147-163.

Chopra, N., J. Lakonishok, and J.R. Ritter, 1992: 'Measuring abnormal Performance - Do Stocks Overreact?' *Journal of Financial Economics* 31, 235-268.

da Costa, N.C.A., 1994: 'Overreaction in the Brazilian Stock Market', *Journal of Banking and Finance* 18, 633-642.

Davis, D.D., and C.A. Holt, 1993: *Experimental Economics*. New Jersey: Princeton University Press.

De Bondt, W.F.M., and R. Thaler, 1985: 'Does the Stock Market Overreact?', *Journal of Finance* 40, 793-805.

De Bondt, W.F.M., and R. Thaler, 1987: 'Further Evidence On Investor Overreaction and Stock Market Seasonality', *Journal of Finance* 42, 557-577.

De Bondt, W.F.M., and R. Thaler, 1989: 'A Mean-Reverting Walk Down Wall Street', *Journal of Economic Perspectives* 3, 189-202.

De Bondt, W.F.M., and R. Thaler, 1990: 'Do Security Analysts Overreact?', *American Economic Review* 80, 52-57.

Edwards, W., 1968: 'Conservatism in Human Information Processing', in *Formal Representation of Human Judgment*, (pp. 17-52), B. Kleinmuntz (ed.). New York,: Wiley.

Fama, E.F., and K.R. French, 1988: 'Permanent and Temporary Components of Stock Prices', *Journal of Political Economy* 96, 246-273.

Forbes, W.P., 1996: 'Picking Winners? A Survey of the Mean Reversion and Overreaction of Stock Prices Literature', *Journal of Economic Surveys* 10, 123-158.

Gamir, E., and Y.Ganzach 1998: 'Overreaction and Underreaction in Analysts' Forecasts', *Journal of Economic Behavior and Organization* 37, 333-347.

Ganguly, A.R., J.H. Kagel, and D.V. Moser, 1998: 'Do Asset Market Prices Reflect Traders' Judgment Biases?', Working Paper, University of Pittsburgh.

Gilovich, T., R. Vallone, and A. Tversky, 1985: 'The Hot Hand In Basketball: On the Misperception of Random Sequences', *Cognitive Psychology* 17, 295-314.

Grether, D.M., 1980: 'Bayes Rule as a Descriptive Model: The Representativeness Heuristic', *Quarterly Journal of Economics* 95, 537-557.

Grether, D.M., 1992: 'Testing Bayes Rule and the Representativeness Heuristic: Some Experimental Evidence', *Journal of Economic Behavior and Organization* 17, 31-57.

Hogarth, R.M., and H.J. Einhorn, 1992: 'Order Effects in Belief Updating: The Belief Adjustment Model', *Cognitive Psychology* 24, 1-55.

McKelvey, R.D., and T. Page, 1990: 'Public and Private Information: An Experimental Study of Information Pooling', *Econometrica* 58, 1321-1339.

Mood, A.M., F.A. Graybill, and D.C. Boes, 1985: *Introduction to the Theory of Statistics*, third edition. Singapore: McGraw-Hill.

Murphy, A.H., and R.L. Winkler, 1970: 'Scoring Rules in Probability Assessment and Evaluation', *Acta Psychologica* 34, 273-286.

Offerman, T., J. Sonnemans, and A. Schram, 1996: 'Value Orientations, Expectations, and Voluntary Contributions in Public Goods', *Economic Journal* 106, 817-845.

Poterba, J.M., and L.H. Summers, 1988: 'Mean Reversion in Stock Prices - Evidence and Implications', *Journal of Financial Economics* 22, 27-59.

Sonnemans, J., and T. Offerman, 1998: 'Is the Quadratic Scoring Rule Behaviorally Incentive Compatible?', Working Paper, University of Amsterdam.

Sonnemans, J., A. Schram, and T. Offerman, 1998: 'Public Good Provision and Public Bad Prevention: the Effect of Framing', *Journal of Economic Behavior and Organization* 34, 143-161.

Stock, D., 1990: 'Winner and Loser Anomalies in the German Stock Market', *Journal of Institutional and Theoretical Economics* 146, 518-529.

Tassoni, C.J., 1996: 'Representativeness in the Market for Bets on National Football League Games' *Journal of Economic Behavior and Organization* 9, 115-124.

Terrell, D., 1994: 'A Test of the Gambler's Fallacy: Evidence from Pari-mutuel Games', *Journal of Risk and Uncertainty* 8, 309-317.

Tversky, A., and D. Kahneman, 1982: 'Representativeness - Belief in the Law of Small Numbers', in *Judgment Under Uncertainty: Heuristics and Biases*, Kahneman, D., P. Slovic and A. Tversky (eds.). Cambridge: Cambridge University Press.

Appendix

				Mean reported probability	
Coin	**Series**	**Type of coin**	**Bayesian probability**	**Exp 1**	**Exp 2**
1	THTHTHHHTTTTTHTTTTTH	fair	23	60.7	45.5
2	HHTTTHHTHHTTTTTHHHTH	fair	40	53.5	43.7
3	TTTTHHTTHHHTTTHTTTHH	false	61	64.6	56.5
4	TTHHHHHHTTTTTTHHTTHH	false	90	82.1	87.1
5	THHTHHHHHHTHTHTTHHTT	fair	11	39.1	28.2
6	TTHTTTHHHHHHHTTTHTHT	fair	40	64.8	66.1
7	TTTTTTTTTTHHHHTHHTHT	false	79	87.6	89.9
8	TTTTTHHTTTHHHHTHHTHH	false	61	55.3	50.2
9	HHTTHTTHTTHHTHHHTHHT	fair	5	22.1	19.0
10	THHHHTTHTHHTTHHHHHTH	fair	23	47.3	42.2
11	HHTHHTTHHHHHHHTTHHTT	fair	61	70.1	53.0
12	TTTHTTHTTTTHTTHTHTTH	fair	5	51.3	38.1
13	HHHHHHTHHHTTTTTTTHTH	false	79	83.5	90.2
14	HHHHHTTTTTHHTHHHHHHH	false	95	83.9	91.6
15	HHTTTHTHHHHTHTTTHHHH	fair	40	40.3	47.3
16	HTHHTTHTHHHTTTTTTHTH	false	11	44.7	30.5
17	TTTTTHHHHTTTTTHHTHTT	false	79	80.5	87.4
18	HTHHHTTTHTHTTHTHHTHH	fair	2	23.3	14.6
19	HHTHTHHHHHHTTHHTHHTT	fair	23	54.6	46.0
20	HTTTTTTHHHHTTTHHHHHT	false	90	82.1	89.7
Total				59.6	55.8