Forthcoming Scandinavian Journal of Economics

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. The recency hypothesis states that traders overweight recent information. Thus, they are too optimistic about winners and too pessimistic about losers. According to the hot hand hypothesis, 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 distinguishing between these hypotheses. The evidence is consistent with the hot hand hypothesis.

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

June, 2004

** University of Amsterdam
CREED / Department of economics
Roetersstraat 11
1018 WB Amsterdam
Netherlands
e-mail: T.J.S.Offerman@uva.nl
e-mail: J.H.Sonnemans@uva.nl

* We would like to thank the editor, three anonymous referees, Uri Gneezy, David Grether, Jan Potters, Arthur Schram, Randolph Sloof and Jörgen Wit for useful comments. Theo Offerman gratefully acknowledges financial support from the Royal Netherlands Academy of Arts and Sciences.

I. Introduction

In the economics profession the proposition that markets yield efficient outcomes has been considered a truism for a long time. The view that markets are efficient has been seriously challenged in recent empirical work on sports markets and financial markets. For both these markets efficiency implies (among others) that the development of the price of a team or 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 drive prices back to intrinsic values.

Gilovich, Vallone and Tversky (1985) were the first to identify a systematic bias in people's beliefs when they predict future sports events. 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. This belief is a bias since they do not find a positive correlation between the actual outcomes of successive shots. Camerer (1989) investigates whether a false belief in autocorrelation is reflected 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. Transaction costs discourage arbitrageurs to exploit the relatively small bias.[1] Overreaction to recent success 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 largely 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

[1]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.

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. They tend to expect more alternations between red and black than a statistician would when red (black) is observed with independent probability 0.5 on each trial (the gambler's fallacy is discussed by Tversky and Kahneman, 1982, and Terrell, 1994). Gamblers falling prey to the gamblers' fallacy as well as bettors misled by the hot hand effect have the tendency to expect too many runs in a series given a certain degree of autocorrelation.

A related phenomenon is reported in financial markets. De Bondt and Thaler (1985) show 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). Usually 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) and Bulkley and Harris (1997) present evidence that even the judgments of professional security analysts reflect systematic biases.[2]

[2]Some studies find overreaction on the long term (3-5 years) and underreaction on the short term (1-12 months) (e.g. Jegadeesh and Titman, 1993, 2001). That is, on the short term recent winners (losers) tend to stay winners (losers). Several attempts have been made to reconcile both phenomena in one model (see Hirshleifer, 2001 for an overview of this literature). For example, the model of Barberis, Shleifer and Vishny (1998) assumes a Bayesian investor who interprets a random walk of earnings as a system which can be in two different regimes: reversal (negative autocorrelation) or trend (positive autocorrelation), with a constant probability of a regime shift. However, a first experimental test of this model by Bloomfield and Hales (2002) fails to show underreaction. Their results show that subjects expect trends which cause overreaction. Other explanations of both under- and

A rational (efficiency preserving) explanation for the phenomenon is provided by risk aversion. If losing firms were the riskier firms, rational traders would need an extra premium in order to buy their stocks. The majority of the aforementioned authors feel that the difference in risk, if it exists at all, is not large enough to explain the difference in expected returns.[3]

Instead, De Bondt and Thaler (1985, 1987) attribute overreaction to the psychological phenomenon of recency. When processing information, people tend to overweight recent information compared with their prior belief or prior data. 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 may thus cause a temporary wedge between stock prices and intrinsic values.

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 similar phenomenon of overreaction is observed. Nevertheless, the phenomenon is explained differently in stock markets than in sports markets. Notice 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, overweight 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 degree of autocorrelation. Assume for

---

overreactions are provided by Daniel, Hirshleifer and Subrahmanyam (1998) and Hong and Stein (1999).

[3]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.

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 there is a trend in the data and that they can detect whether a firm is in a good or a bad shape. As a consequence, they would put too much value on winning firms and too little value on losing firms. The 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.

Although the observed phenomenon of overreaction is the same in stock markets and sports markets, there is an important difference between the two markets. In sports markets, bettors bet on a team and their payoff is simply determined by the actual performance of the team in the subsequent sports contest. In stock markets, traders form beliefs about the future prices of the stocks. Here, unlike in sports markets, the uncertainty about the intrinsic value of the stock is rarely resolved. As a consequence, most traders buy stocks in an attempt to resell them later at a higher price. In this sense, the average opinion of the other traders is at least important as the intrinsic value of the stock. In stock markets, there is room for traders to rationally ride a bubble and to exploit misguided beliefs of other traders. This speculation motive is absent in sports markets, where traders do not buy bets in an attempt to resell them at a higher price.

The main goal of this paper is to determine in an experiment whether recency or hot hand is the better explanation of overreaction. We design an experiment that allows us to distinguish between the two explanations. Subjects are rewarded for the accuracy of their beliefs about an uncertain event that is resolved after the predictions have been made. Like in sports markets, there is no speculation motive in our experiment because there is no possibility to buy bets and resell them at a higher price. So our results are clearly most relevant for the literature on sports markets. We do believe, however, that they also provide a useful direction for the discussion about the cause of overreaction in stock markets.

Consider the following experimental setup. A coin is randomly selected 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). Notice that in the long run a false coin will produce an equal number of heads and tails. So in this respect our terminology differs from everyday language, where false coins tend to refer to coins that are loaded to either side. In our experiment, 'unfair' solely refers to autocorrelation.

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,

$$B = \frac{P\ [false\,coin] * P\ [data|\ false\ coin]}{P\ [false\,coin] * P\ [data|\ false\ coin] + P\ [fair\ coin] * P\ [data|\ fair\ coin]} \tag{1}$$

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

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. Likewise, a series produced by a false coin with positive autocorrelation will 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.[4] Recency occurs if a decision-maker overweights the recent information contained in the number of alternations in the series of coin tosses compared to the prior information[5] 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.[6] 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 underweight new evidence compared with the prior belief.

---

[4]Notice that we propose a narrow definition of the hot hand effect, that makes the interpretation of the statistical analysis easier. The hot hand effect is also associated with people's poor understanding that fair coins can generate long *streaks*. At the end of section 3, we report how streaks affect subjects judgments in the current setting.

[5]Our interpretation of recency is slightly different from its normal use in psychology. There, the term recency generally refers to weighting recent observations more than earlies ones. A direct consequence of recency is that prior information does not receive sufficient weight in the belief formation process.

[6]Grether (1980) 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.

Such a decision-maker would be affected by conservatism, and make errors in the opposite direction of the errors implied by recency.[7] The design makes it possible to discriminate between these four hypotheses. Together 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 II describes the experimental design in more detail. Section III provides the results for experiment 1 and 2. Section IV contains a concluding discussion.

II. Experimental design

[7]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 overweight prior information when identifying aircrafts as friendly, hostile or unknown.

Both the instructions and the decision phase of the experiment are computerized.[8] 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 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 subjects receive information about how they can 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

[8]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.

Dutch guilder).

The quadratic scoring rule is an incentive compatible mechanism for expected value maximizers. McKelvey and Page (1990) use the quadratic scoring rule 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 a subject's best interest to report beliefs truthfully. 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.[9]

Experiment 2 is designed to investigate whether the biases from Bayesian updating are robust

[9] 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.

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 in a range of 15.50 to 21.25 guilders was earned by subjects in about 45 minutes.[10] 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 in a range of 18.00 to 22.10 guilders was earned by subjects in experiment 2.

[10] One guilder can be exchanged for approximately 0.45 euro.

III. Results

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.

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.

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).

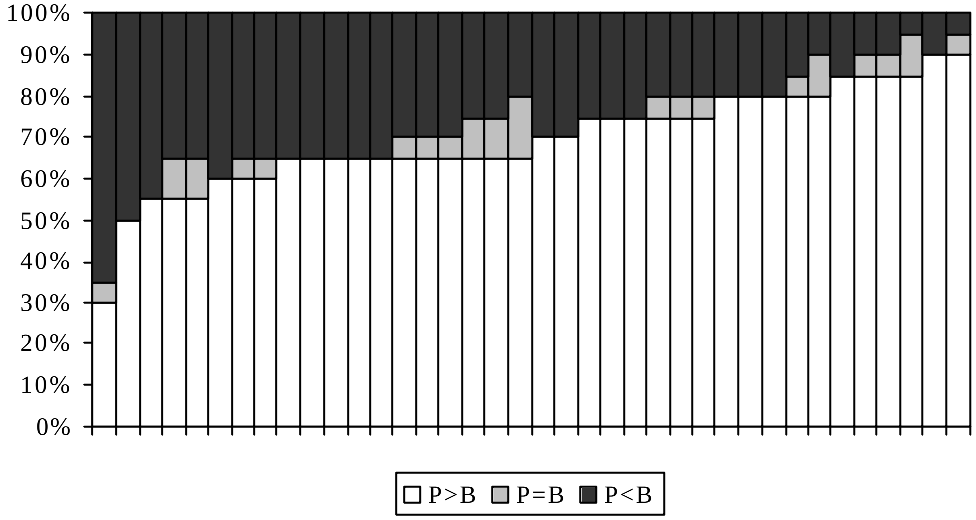

FIGURE 1: biases per subject in experiment 1. Each of the 38 subjects is represented by a bar. The upper (lower) part of the bar displays the percentage of coins for which a subject underestimated (overestimated) the probability that it was false; the middle part represents the percentage of accurate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

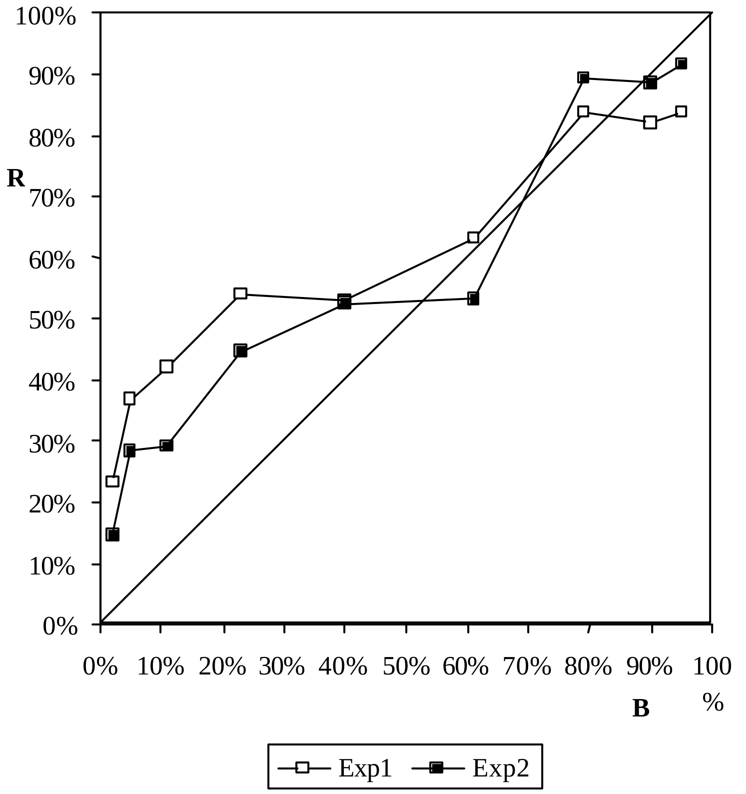

FIGURE 2: mean reported probabilities (R: vertical axis) in experiments 1 and 2 as function of the Bayesian probabilities (B: horizontal axis). The straight line R=B is added as a benchmark.

A pair-wise non-parametric comparison of the four hypotheses is made to assess which hypothesis takes best account of the biases in subjects' judgments. Table 2 summarizes the results. For some comparisons of hypotheses we focus on a subset of the coins. 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.

TABLE 2
Non-parametric 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 shows 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).

Next we deal in a parametric way with the question which hypothesis takes best account of the data. The evidence in Figure 2 supporting the hot hand effect is more clearcut for coins that seem fair than for coins that seem false. This might be explained by the presence of random error: it is more likely that the positive bias caused by a hot hand effect is offset by a negative bias caused by random error 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, we estimate some models using maximum likelihood techniques. The Bayesian model maintains the hypothesis that subjects are Bayesian and that all errors are random error:

$$R = B + e\ . \tag{2}$$

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, σ^2) distribution. The distribution of the error terms is truncated, because subjects never report probabilities smaller than 0% or greater than 100%.

We compare the Bayesian model with a more general model that allows subjects to deviate both along the conservatism versus recency dimension and the hot hand versus cold hand dimension. The following flexible specification is inspired by the work of Edwards (1968).

$$R = \frac{a\left(\frac{B}{1-B}\right)^{b}}{1+a\left(\frac{B}{1-B}\right)^{b}} + e \tag{3}$$

The parameter α captures deviations along the hot hand versus cold hand dimension. The case where α >1 implies that subjects overestimate the probability of a false coin for all coins, as expected on the basis of the hot hand hypothesis. With α <1 the equation implies that subjects always underestimate the probability of a false coin, as expected by the cold hand hypothesis. The parameter β captures deviations along the conservatism versus recency dimension. If β is set smaller than 1 in the equation, subjects' predictions are biased towards 50%, in accordance with conservatism. With a β larger than 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. Figures 3 and 4 illustrate how changes in the parameters affect the model's predictions compared to the Bayesian benchmark.

First we assume that all subjects use the same rule when forming their judgments. Besides the general model described by (3) allowing for biases in two dimensions, we also estimate a model where subjects only fall prey to a bias in the conservatism versus recency dimension (by setting α =1), and a model where subjects only fall prey to errors in the hot hand versus cold hand dimension (by setting β=1). Notice that the Bayesian model is nested in the general model described by equation (3): by setting the parameters α and β both equal to 1, the model simplifies to the Bayesian model. We add the results for the random model as a lower bound for the expected performance of the model. According to the random model each of the 101 possible percentages from the set {0, 1, …, 99, 100} is selected with equal probability. The random model is nested in the Bayesian model, by letting $\sigma \to\infty$.

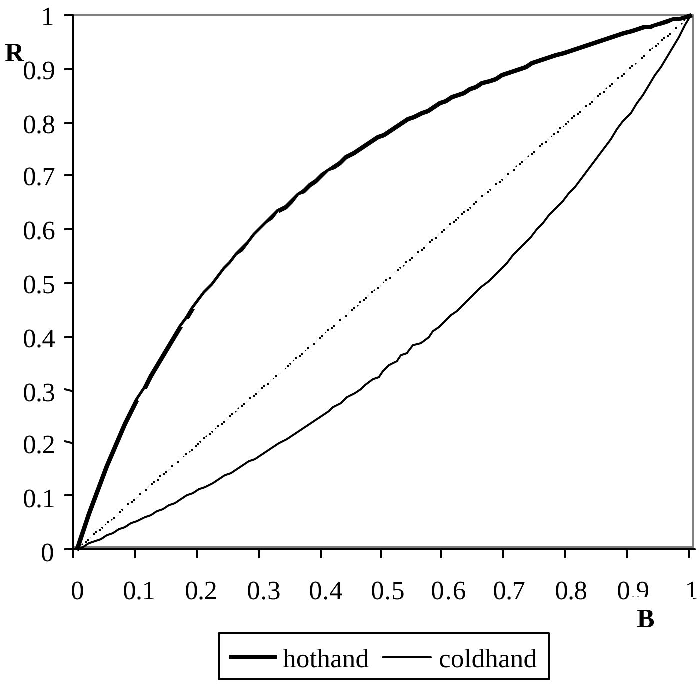

FIGURE 3: The model of equation (3) with α =3.5; β =1 (hothand) and α =0.5; β =1 (coldhand). The straight line R=B is added as a benchmark.

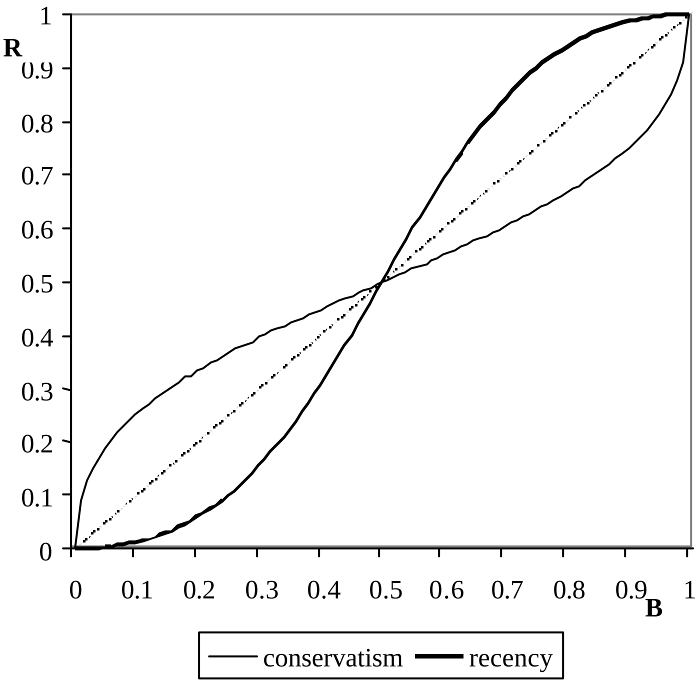

FIGURE 4: The model of equation (3) with α =1; β =0.5 (conservatism) and α =1; β =2 (recency). The straight line R=B is added as a benchmark.

Table 3 summarizes the results of this estimation procedure. The Bayesian model explains the data significantly better than the random model. Of the two models that introduce restrictions into the general model, the model that allows subjects to deviate along the hot hand versus cold hand dimension provides the strongest improvement in likelihood compared with the Bayesian model. The parameter a is estimated to be substantially larger than 1, identifying a clear hot hand effect. The most general model without restrictions does not improve the fit of the data significantly compared with this model. It is concluded that hot hand plus random error provides the best explanation of the data.

TABLE 3
Maximum likelihood results assuming homogeneity

model	*restrictions*	*estimated bias parameter(s)*	*s*	*-logL*
(i) random	α=1; β =1; $\sigma \rightarrow \infty$	--	--	3507.5
(ii) Bayes	α =1; β =1	--	37.1	3388.2**
(iii) only hot hand/ cold hand	β =1	α =3.6	31.2	3312.3**
(iv) only recency/ conservatism	α =1	β =1.6	40.3	3384.5*
(v) general, both biases	--	α =3.1 β =0.9	29.8	3311.1

Notes: The loglikelihood is computed on the basis of 760 observations. * (**) Indicates significance of the likelihood ratio test at the 5% (1%), when the likelihood of the model and the nested model with one parameter less is compared.

The foregoing analysis makes the rather strong assumption that all subjects use the same model when forming their judgments. Now we relax this assumption by allowing subjects to use different models. However, it is still assumed that a particular subject makes all 20 choices according to the same model. Models 1 to 5 will refer to the hot hand model (α >1; β =1), the cold hand model (α <1; β =1), the conservatism model (α =1; β <1), the recency model (α =1; β >1) and the Bayesian model (α = β =1), respectively. P_j denotes the probability that an arbitrary subject uses model j. We define $P_1+...P_5=1$. If a subject uses model j, we assume that the error component is drawn from a truncated N(0, σ_j^2) distribution. Let $R_{i,t}$ denote the report of subject i for coin t. Then, the unconditional likelihood function of subject i's reports for coins 1-20 $L(R_{i,1},..,R_{i,20})$ is given by:

$$L(R_{i,1},...,R_{i,20}) = \sum_{j=1}^{5} [P_j * \prod_{t=1}^{20} L(R_{i,t} \mid model_j)] , \qquad (4)$$

where $L(R_{i,t}|model_j)$ denotes the conditional probability that subject i reports $R_{i,t}$ for coin t when (s)he uses model j. The total unconditional likelihood function for the hybrid model encompassing models 1 to 5 is then obtained by multiplying all subjects' unconditional likelihood functions. Table 4 reports the estimates of the parameters that maximize this total likelihood function.

TABLE 4
Maximum likelihood results allowing heterogeneity

hybrid model	*Probability*	*bias parameter*	*s*	*-logL*
1. hot hand	0.86	α =3.7	27.5	3287.7
2. cold hand	0.00	α undetermined	σ undetermined	
3. conservatism	0.03	β =0.01	11.4	
4. recency	0.00	β undetermined	σ undetermined	
5. Bayes	0.11	--	69.4	

Notes: The loglikelihood is computed on the basis of 760 observations. The column "probability" reports the estimate for the proportion of the population using the particular model. (Only four probabilities were estimated, and the fifth followed from the definition that $p_1+...+p_5=1$.) When a probability is estimated to be equal to 0, the corresponding bias parameter cannot be determined.

A remarkable result is that for the present updating task no subject seems to use the rules based on recency and cold hand. A large majority of 86% of our subjects is estimated to use the hot hand model, 11% is estimated to act in accordance with the Bayesian model (but in a noisy way), and the remainder 3% is estimated to be conservative updaters. The addition of the Bayesian and conservative types to the hot hand model (iii) reported in Table 3 leads to a significant increase of the likelihood of the data (Likelihood ratio test at the 1% level).[11]

One may wonder whether other factors or biases exist that also play a role in determining a

[11] We have chosen to introduce heterogeneity in a parsimonious way by restricting the number of possible types. An alternative procedure is to estimate equation 3 for each subject separately. Qualitatively this leads to the same results. The main difference is that the Bayesian model is not rejected for a larger part of the population. This difference is not surprising because we have relatively few observations per subject. The results of this alternative classification method are: 6 subjects (15.8%) have both a significant hot hand and a significant conservatism parameter; 20 subjects (52.6%) only have a significant hot hand parameter; 1 subject (2.6%) only has a significant conservatism parameter; for 9 subjects (23.6%) the Bayesian model cannot be rejected and for the remaining 2 subjects (5.2%) the random model cannot be rejected.

subject's reported probability of a false coin. We consider two possibilities: (i) the posterior subjective probability may be affected by a bias in the relative frequency of heads and tails in the series; (ii) the posterior subjective probability of a false coin may be affected by the length of the maximal streak in the series generated by the coin.

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.

The other factor considered is the length of the maximal streak in the series. A Bayesian observer should allocate equal probabilities to coins with equal numbers of alternations (*cf*. equation 1). In contrast, subjects tend to conclude that a series provides stronger evidence for a false coin if the length of the maximal streak is longer. 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.

This result questions the descriptive validity of the Bayesian model. Although the Bayesian framework provides a parsimonious and useful way to separate between different hypotheses, it may not be an accurate model of how people actually form beliefs. In future research it will be interesting to develop an alternative descriptive model of expectation formation and to test it in a larger set of experiments where the prior probabilities of the coins are varied. These data will be useful to acquire a better understanding about how posterior probabilities vary with prior probabilities. These additional data will also allow to test for the robustness of the result that subjects' belief formation process is driven by the hot hand effect.

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.

IV. 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 error part. Recency cannot explain our data.

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. Besides the recency effect (or insensitivity to prior probabilities) and the hot hand effect, representativeness also predicts judgments to be insensitive to sample size. Rabin (2002) discusses some interesting economic implications of the "law of small numbers". For instance, consider the case where a mutual fund manager judges the performance of financial analysts. If, in truth, all analysts are average, a believer in the law of small numbers may quickly fire young analysts who accidentally perform below average and keep the ones that perform above average. After a short while, the analysts who performed above average will regress towards the mean, and the manager may correct her positive view of these analysts, but not her negative view of the fired analysts. Her overall view of the quality will therefore be biased downwards.

In the present setting, the hot hand effect appears to be a stronger force than the countervailing force of recency. Here, subjects also seem to believe that the evidence for the falseness of a coin is stronger if the length of the maximal streak of its series is longer. These findings may improve our understanding of the concept 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, 2000).

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 future 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 overweights 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 overweighting recent information.

References

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

Adelman, L., Bresnick, T., Black, P.K., Marvin, F.F. and Sak, S.G. (1996), Research with Patriot Air Defense Officers: Examining Information Order Effects, *Human Factors* 38, 250-261.

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

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

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

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

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

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

Bulkley, G. and Harris, R.D.F. (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., Loewenstein, G. and Weber, M. (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., Lakonishok, J. and Ritter, J.R. (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.

Daniel, K., Hirshleifer D. and Subramanyam, A. (1998), Investor Psychology and Security Market under- and overreaction, *Journal of Finance* 53, 1839-1885.

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

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

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

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

De Bondt, W.F.M. and Thaler, R. (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 French, K.R. (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.

Ganguly, A.R., Kagel, J.H. and Moser, D.V. (2000), Do Asset Market Prices Reflect Traders' Judgment Biases?, *Journal of Risk and Uncertainty* 20, 219-45

Gilovich, T., Vallone, R. and Tversky, A. (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.

Hirshleifer, D. (2001), Investor Psychology and Asset Pricing, *Journal of Finance* 56, 1533-1597.

Hong, H. and Stein, J.C. (1999), A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, *Journal of Finance* 54, 2143-2184.

Jegadeesh, N. and Titman, S. (1993), Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65-91.

Jegadeesh, N. and Titman, S. (2001), Profitability of Momentum Strategies: An Evaluation of Alternative Explanations, *Journal of Finance* 56, 699-720.

McKelvey, R.D. and Page, T. (1990), Public and Private Information: An Experimental Study of Information Pooling, *Econometrica* 58, 1321-1339.

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

Offerman, T., Sonnemans, J. and Schram, A. (1996), Value Orientations, Expectations, and Voluntary Contributions in Public Goods, *Economic Journal* 106, 817-845.

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

Rabin, M. (2002), Inference by Believers in the Law of Small Numbers, *Quarterly Journal of Economics* 117, 775-816.

Sonnemans, J., Schram, A. and Offerman, T. (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 Kahneman, D. (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

Coin	Series	Type of coin	Bayesian probability	Mean reported probability Exp 1	Mean reported probability 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
Average			45.9	59.6	55.8