# The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?

PETER AYTON
*City University, London, England*

and

ILAN FISCHER
*University of Haifa, Haifa, Israel*

The representativeness heuristic has been invoked to explain two opposing expectations—that random sequences will exhibit positive recency (the hot hand fallacy) and that they will exhibit negative recency (the gambler's fallacy). We propose alternative accounts for these two expectations: (1) The hot hand fallacy arises from the experience of characteristic positive recency in serial fluctuations in human performance. (2) The gambler's fallacy results from the experience of characteristic negative recency in sequences of natural events, akin to sampling without replacement. Experiment 1 demonstrates negative recency in subjects' expectations for random binary outcomes from a roulette game, *simultaneously* with positive recency in expectations for another statistically identical sequence—the successes and failures of their predictions for the random outcomes. These findings fit our proposal but are problematic for the representativeness account. Experiment 2 demonstrates that sequence recency influences attributions that human performance or chance generated the sequence.

*It is principally at games of chance that a multitude of illusions support hope and sustain it against unfavourable chances.* (Laplace, 1796/1951)

The idea that beliefs about probability show systematic biases is somewhat older than experimental psychology. Throughout his "Essai Philosophique sur les Probabilités," Laplace (1796) was concerned with errors of judgment and even included a chapter concerning "illusions in the estimation of probabilities." It is here that we find the first published account of what is now widely known as the *gambler's fallacy*—the belief that, for random events, runs of a particular outcome (e.g., heads on the toss of a coin) will be balanced by a tendency for the opposite outcome (e.g., tails).

Early experiments in probability learning empirically confirmed the reality of this bias in tasks where subjects were asked to predict the next outcome in a series of random binary alternatives. Although Jarvik (1951) found evidence for a tendency to predict the same as the last event (positive recency), the chance of predicting the same as the past diminished after a run of two such events transforming into a tendency to predict the opposite of the last event (negative recency), for even longer runs, this tendency intensified.

Numerous early studies confirmed this general pattern of findings (for a review, see Lee, 1971, chapter 6). Several studies found that, in experiments with 1,000 or more trials, recency effects disappear (e.g., Derks, 1962, 1963; Edwards, 1961). Witte (1964) showed that negative recency in subjects' responses could last for several days' sessions, including 900 trials, but that eventually it appeared to diminish. Negative recency is also commonly observed (although not always, see Budescu, 1987) in the responses of subjects who are asked to generate or identify random sequences (see, e.g., Bar-Hillel & Wagenaar, 1991).

Estes (1964) suggested that the negative recency observed in these experiments was a habit learned from life that was revealed in the laboratory and, in longer experiments, was gradually extinguished. Yet, if the gambler's fallacy is a habit learned in everyday life, where might it be learned? Plainly, it is an inappropriate response to situations where there is conditional independence between the successive outcomes of a random process. Nonetheless, outside of gambling casinos and psychology laboratories, there are few—if any—circumstances where one can safely assume conditional independence of a succession of events. Perhaps, then, the gambler's fallacy reflects adaptation to uncertain situations where negative recency

is exhibited. One obvious candidate is where a finite population of outcomes is sampled without replacement. Under these circumstances, expectations with negative recency have some validity because observing a particular outcome lowers the chances of observing that outcome the next time. Accordingly, a number of authors have suggested that the experience of negative recency in life might be responsible for the gambler's fallacy in experimental tasks where subjects are asked to generate or recognize random sequences (Ayton, Hunt, & Wright, 1989, 1991; Lopes, 1982; Lopes & Oden, 1987; Neuringer, 1989; Triesman & Faulkner, 1990). For example, Pinker (1997) is critical of the presumption of faulty reasoning typically accompanying observations of the gambler's fallacy:

> It would not surprise me if a week of clouds really did predict that the trailing edge was near and the sun was about to be unmasked, just as the hundredth rail road car on a passing train portends the caboose with greater likelihood than the third car. Many events work like that. . . . An astute observer should commit the gambler's fallacy. A gambling device is by definition a machine designed to defeat our intuitive predictions. It's like calling our hands badly designed because they fail to get out of handcuffs. (p. 346)

A rather different account of the gambler's fallacy was offered by Kahneman and Tversky (1972), who presented a cognitive explanation of the gambler's fallacy in terms of the operation of the representativeness heuristic. They argued that people expect the essential characteristics of a chance process to be represented not only globally in an entire sequence of random outcomes but also locally in each of its parts. Thus, despite their statistical inevitability, long runs of the same outcome lack local representativeness and are thereby not perceived as representative of the expected output of a random device. Consequently, subjects will expect runs of the same outcome to be less likely than they are.

However, the idea that perceptions of randomness are governed by representativeness has also been used to explain the exact opposite phenomenon—that in a random sequence, people have an incorrect expectation that a run of the same outcome will continue. The "hot hand" fallacy was observed by Gilovich, Vallone, and Tversky (1985), who noted that most people associated with the game of basketball believe that a player who has just scored several times in a row is now more likely to score—because he or she is "hot." However, when these authors computed the sequential dependencies between successive scoring attempts of players, they found that there was no such dependency; indeed, if anything, players who have had a run of successful scoring attempts are somewhat *less* likely to score next time. Gilovich et al. explained that judgment by representativeness can lead people to reject the randomness of sequences that contain the expected number of runs because the appearance of long runs in short samples makes the sequence appear unrepresentative of randomness (see also Gilovich, 1991; Tversky & Gilovich, 1989a). Thus, the representativeness heuristic has been used to explain the expectation that a random sequence will exhibit both negative and positive recency.

Such an account for the hot hand fallacy raises a question regarding the gambler's fallacy: If observation of the runs associated with conditional independence in basketball is reason for observers to reject the notion that the sequence of success and failure is random, why do people not come to a similar conclusion in situations where the gambler's fallacy has been observed? Why, for example, do roulette players, on observing the distribution of runs of outcomes from the wheel, not reject the hypothesis that the wheel is random and believe that the wheel gets hot—in the sense that runs of red make it *more* likely for red next time? Given that different expectations arise for essentially the same patterns of data, one possibility is that differences in people's *prior* expectations for different processes have some effect.

It is fairly easy to imagine credible reasons to support the expectation that successive attempts at scoring in basketball will be positively autocorrelated. For example, plausibly, a series of attempts at scoring might be affected positively by success and resulting increases in confidence and/or negatively by failure and resulting decreases in confidence. Or there might be a fatigue and recovery cycle underlying repeated responses. In either case, positive recency would result. Indeed, Gilden and Wilson (1995, 1996) have shown that for golf putting, dart throwing, and auditory and visual signal detection, there *are* streaks in performance; Adams (1995) reported "momentum" in the performance of pocket billiards players, Dorsey-Palmateer and Smith (2004) found evidence for streaks in ten pin bowlers, and Smith (2003) reported that horseshoe pitchers have modest hot and cold spells. Accordingly, belief in the hot hand is not always fallacious. Perhaps then, elaborating on Estes's argument, people have learned to expect the hot hand from observing human performances where it occurs. People seem to believe that people can get "hot" (exhibit positive recency), but that inanimate devices cannot.

For chance events such as coin tosses or spins of a roulette wheel, where concepts of fatigue, confidence, or motivation do not pertain, it is harder to envisage reasons why sequences should be positively autocorrelated. For such events, people typically behave as if expecting a process of sampling without replacement. Indeed, our quote from Pinker (1997) gives two examples where sequences produced by inanimate processes would be expected to produce negative recency.

In our first experiment, we sought to investigate the possibility that incorrect expectations of positive and negative recency in random sequences arise as a function of the category of event that subjects experienced. Where subjects construe a sequence of outcomes as reflecting human performance, we hypothesize positive recency. But where subjects expect that outcomes are due to an inanimate mechanism, we hypothesize negative recency.

Our second experiment adopts a different strategy and utilizes a different task to explore the same notion. In-

stead of studying the expectations of recency that people make when presented with sequences with different sources of variance, we investigate the inferences that people make about the sources of sequential variance when presented with sequences with different properties. We present sequences of outcomes with varying levels of positive and negative recency and measure the attributions that are made about the likely source of the sequence. We hypothesize that sequences with positive recency will be attributed to human skilled performance, whereas those with negative recency will be attributed to inanimate chance mechanisms.

## EXPERIMENT 1

In our first study, we asked subjects to predict the next outcome in a random series and then measured sequential dependencies in these responses and, separately, subjects' ex ante confidence in the success of each prediction. We hypothesize that while their predictions will show negative recency with respect to the outcomes (the gambler's fallacy), their beliefs in the success of their predictions will show positive recency with respect to the runs of success and failure of their predictions (the hot hand fallacy).

We explored these different expectations in an experiment where subjects were either instructed to forecast the next event in a binary time series (red or blue), or to make bets in a simple binary version of roulette. We studied this instructional variable because the probability learning literature contains some (inconsistent) evidence that it alters behavior (Lee, 1971).

Having made their predictions or bets, all subjects adjusted a scale to indicate how confident they were in their bets/predictions. They then played computer roulette and received feedback. The outcomes on all trials were randomly determined.

### Method

**Subjects**. Thirty-two undergraduate students at City University participated in the experiment.

**Stimuli and Materials**. Computer software simulated the action of a schematic roulette wheel and provided the tools to elicit confidence measures and record subjects' responses. A prediction panel—initially colored white—could be changed to red or blue to indicate predictions, which could be altered until finally confirmed by pressing the "enter" key.

A rectangular frame served to elicit confidence. The left end of the frame was labeled with the words "no confidence," and the right end with the words "strong confidence." No other labels or numerical values were indicated. The subjects signaled higher levels of confidence by pressing the right arrow key, causing the frame to change color from black to yellow, starting at the "no confidence" edge and moving toward the "high confidence" edge; the left arrow key cleared the yellow color in the opposite direction. This enabled a resolution of 21 units, mapped by the software along the range 0–100 (i.e., 0, 5, 10, . . . , 90, 95, 100).

A simplified roulette wheel—the upper half of the wheel was colored red and the lower half blue—was displayed. A black "clock-hand" initially pointed left on the borderline between the red and blue segments. When triggered, the hand spun around several times and stopped to point at either the red or the blue color. Final stopping locations were selected randomly by the computer with an equal probability of .5 for each color. The computer's random number generator was seeded with a new initial value (derived from the computer's clock) for each trial.

**Procedure**. The subjects were randomly assigned to either the "gambling" or "forecasting" instructional condition; but because not all recruited subjects attended, 15 were assigned to the forecasting condition and 17 to the gambling condition. The subjects in the gambling condition were told that their task was to gamble on the computerized roulette outcome, whereas the subjects in the forecasting condition were asked to forecast the computer's programmed algorithm outcome. The subjects in the latter group were told that "there is a system which determines the outcome of the wheel; it is possible to learn the system so as to do better than chance"). For both groups, the actual roulette outcomes were equally likely and randomly drawn. The subjects were promised a bonus of 5 pennies for each correct prediction and a deduction of 5 pennies for each wrong prediction. They were guaranteed a minimum of 3 pounds for participation, even if they made a loss during the experiment.

On each trial, the subjects predicted red or blue and then indicated their confidence in their prediction. The roulette wheel was spun, and an outcome panel appeared red or blue, accordingly. Simultaneously, a tone sequence played to increase the salience of winning and losing; successful predictions were followed by a triumphant fanfare and failed predictions by a grim dirge.

After 200 trials, a summary of the accumulated wins and losses appeared on the screen. The subjects were then debriefed and paid.

### Results

**Negative recency in expectations of outcomes (the gambler's fallacy)**. We collated the data so we could examine the extent to which responses in choosing red or blue were sequentially dependent on previous outcomes of red or blue. Figure 1 plots the probability of predicting (for forecasters and gamblers) the same color as the last outcome as a function of the number of consecutive times that outcome had just occurred. A 2 (gambling/forecasting) $\times$ 5 (run length) analysis of variance (ANOVA) revealed that run length of previous outcomes was a significant determinant of the tendency to predict the same outcome as before [$F(4,120) = 5.11$, $MS_e = 240.99$, $p = .001$] and that there was a significant linear trend by run length [$F(1,30) = 8.69$, $MS_e = 518.38$, $p = .006$]. The longer the run of a particular color, the less likely subjects were to predict that color the next time. Thus, negative recency—the gambler's fallacy—is apparent in the choices of the subjects. There was no effect of task (forecasting vs. gambling) [$F(1,30) = 0.04$, $p = .844$] and no significant interaction between task and run length [$F(4,120) = 1.25$, $p = .29$].

**Positive recency in expectations of success and failure (the hot hand fallacy)**. To examine the data for the presence of a hot hand effect, we collated subjects' trial-by-trial confidence ratings to see how they varied with runs of success and failure in the betting/predicting task. One subject (a gambler) did not experience any runs of success 5 events long, so a value was estimated by calculating the mean confidence the other 16 gamblers gave to this event and substituted into the analysis.
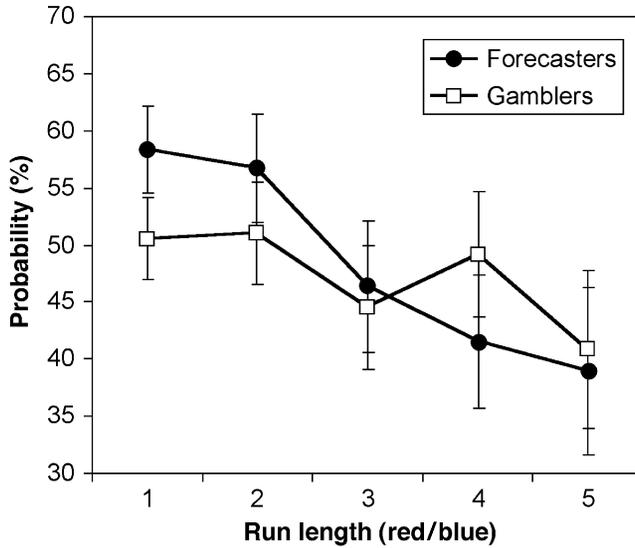
Figure 1. Percent probability (±*SE*) of predicting the same color as the last outcome.

A 2 (task) × 5 (run length) ANOVA on the confidence ratings shows that run length (of successful past predictions) was a significant factor affecting confidence [$F(4,120) = 20.25$, $MS_e = 4.97$, $p < .001$]. Figure 2 shows that the confidence expressed by subjects increases according to the runs of success that they obtained in the task. There was a significant linear trend in confidence as a function of run length of successes [$F(1,30) = 14.70$, $MS_e = 12.98$, $p < .001$]; subjects experiencing a run of successful predictions increased their confidence in their next prediction. The ANOVA also revealed that gamblers were generally more confident than forecasters [$F(1,30) = 65.83$, $MS_e = 48.55$, $p < .001$], perhaps because the forecasters, told that outcomes were somewhat predictable,

judged from a different standpoint: They may have been dismayed by being unable to achieve higher expectations of success. This may have reduced their confidence relative to the gamblers, who had no reason to expect better than chance performance.

There was also an interaction effect between run length and task [$F(4,120) = 12.82$, $MS_e = 4.97$, $p < .001$]. Whereas both forecasters and gamblers showed an increase in confidence over runs of success, the pattern for each is somewhat different. Whereas there is a statistically significant positive linear trend across run lengths of 1–5 for gamblers [$t(16) = 3.79$, $p < .01$], the downturn in confidence for run lengths of 5 for forecasters prevents the corresponding trend from achieving statistical significance [$t(14) = 1.60$, $p = .13$]. Nevertheless, across run lengths of 1–4 there is a significant positive linear trend for both forecasters [$t(14) = 6.46$, $p < .0001$] and gamblers [$t(16) = 2.97$, $p < .01$].

We also collated the data so as to examine the effects on confidence in predictions as a function of runs of failure. Figure 3 shows that, as runs of failed predictions were experienced, subjects' confidence in their predictions generally decreased.

A 2 (task) × 5 (run length) ANOVA confirmed that these effects were statistically significant. Run length of failed predictions had a significant effect on subjects' confidence [$F(4,120) = 28.23$, $MS_e = 6.67$, $p < .0001$]. There was also a significant negative linear trend [$F(1,30) = 29.76$, $MS_e = 10.87$, $p < .0001$] by run length, confirming that the longer the run of failure, the less confident subjects were in their predictions. Task was also a significant factor; forecasters were again generally less confident than gamblers [$F(1,30) = 13.82$, $MS_e = 95.57$, $p = .001$]. The interaction between task and run length was also significant [$F(4,120) = 16.77$,
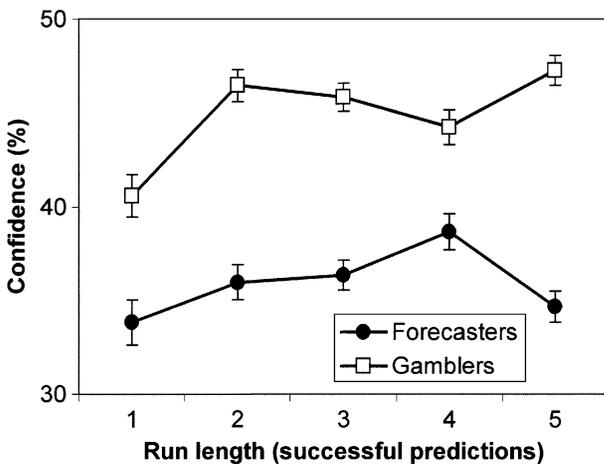


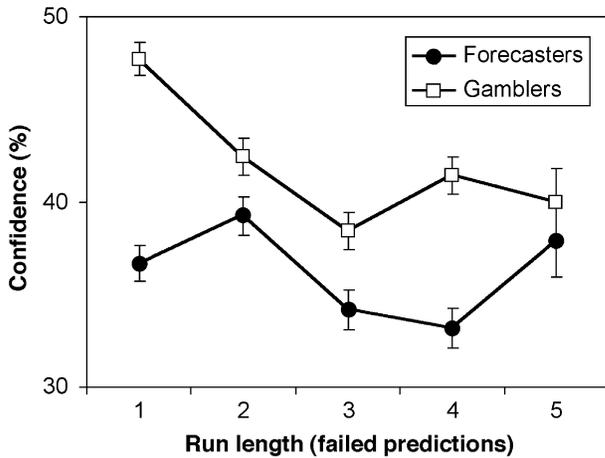Figure 2. Percent confidence (±*SE*) in predicting/gambling as a function of runs of success.

Figure 3. Percent confidence (±$SE$) in predicting/gambling as a function of runs of failure.

$MS_e = 6.67$, $p < .0001$]. While both forecasters and gamblers showed a decrease in confidence over the runs of failure, the precise pattern for each is again somewhat different. Whereas for run lengths of 1–5, there is a statistically significant negative linear trend for gamblers [$t(16) = -9.21$, $p < .001$], for forecasters the upturn in confidence after five failures renders the corresponding linear trend not statistically significant [$t(14) = -1.07$, $p = .3$]. Nevertheless, across run lengths of 1–4 there is a significant negative linear trend for both forecasters [$t(14) = -6.52$, $p < .0001$] and gamblers [$t(16) = -12.52$, $p < .0001$].

The analysis of the link between confidence and runs of success and failure reveals that subjects increase their prospective confidence in their choices as a function of runs of successful predictions and decrease confidence as a function of failed predictions. This applies for both forecasters and gamblers up to run lengths of 4. Thus, in contrast to their predictions of the color of the outcome of the roulette wheel, subjects' changing confidence presumes that their sequence of winning and losing will exhibit positive recency. Subjects apparently believed that they got hot or cold, although there were no serial dependencies in the outcomes that they were predicting.

**Confidence in predictions with negative and positive recency**. Further illustration of the two effects studied here was revealed when we compared subjects' confidence on the occasions when they chose to predict that the wheel would repeat the last outcome of red or blue or alternate from its last outcome. We collated the confidence data according to whether subjects predicted that the wheel would repeat or alternate its last response as a function of the run lengths of red or blue that had occurred. Treating the choice of alternate or repeat as two levels of an independent variable, we performed a 2 (choice) × 2 (task) × 5 (run length) ANOVA on the confidence data. This analysis revealed a significant inter-

action between choice and run length [$F(4,120) = 2.61$, $MS_e = 111.26$, $p = .039$] plotted in Figure 4.

The interaction reveals that when subjects chose to predict that the wheel would repeat its last outcome, they became less confident as the run became longer. However, if they predicted that the outcome of the wheel would be different from the last time, they became more confident as the run length increased, although there is a noticeable decrease in confidence after 5 runs of the same outcome. Linear trend analyses revealed that for predictions the same as the last outcome, there is a significant negative linear trend in confidence by run length [$t(31) = -2.30$, $p < .05$]. For predictions that the next outcome would be different, the positive trend is statistically significant across run lengths of 1–4 [$t(31) = 1.82$, $p < .05$, one-tailed test], although it is not statistically significant across run lengths of 1–5 [$t(31) = 1.03$, $p = .31$]. Thus, subjects' confidence in their own predictions varies according to whether they predict with negative or positive recency; predictions consistent with the gambler's fallacy are more confident, and (at least up to run lengths of 4) increasingly so with run length, whereas predictions inconsistent with the gambler's fallacy are less confident and become even less so with run length.

## Discussion

The results of the experiment confirm our hypotheses; while subjects' predictions show negative recency with respect to the sequence of outcomes of the roulette wheel (the gambler's fallacy), their beliefs in the sequence of success and failure of their predictions show positive recency (the hot hand fallacy). Note that, in terms of their statistical characteristics, the sequence of outcomes of the roulette wheel (red and blue) and the sequence of outcomes of the predictions of the subjects (win and lose) are indistinguishable; they are each binary random processes with a probability of .5 for each event. Nonethe-
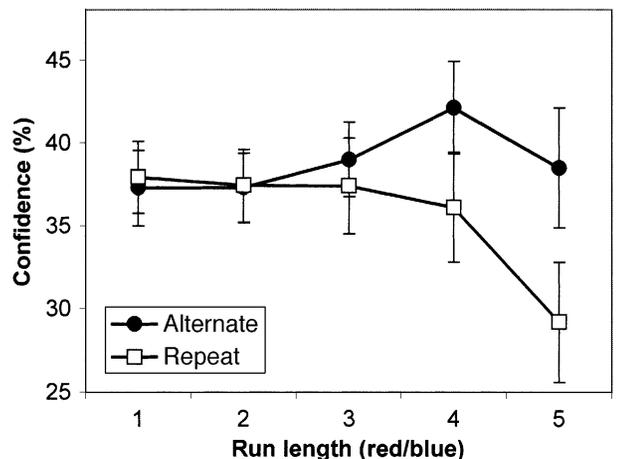


Figure 4. Percent confidence (±$SE$) as a function of predicting repeating/alternating runs of red/blue.

### A     Basketball or Coin Toss



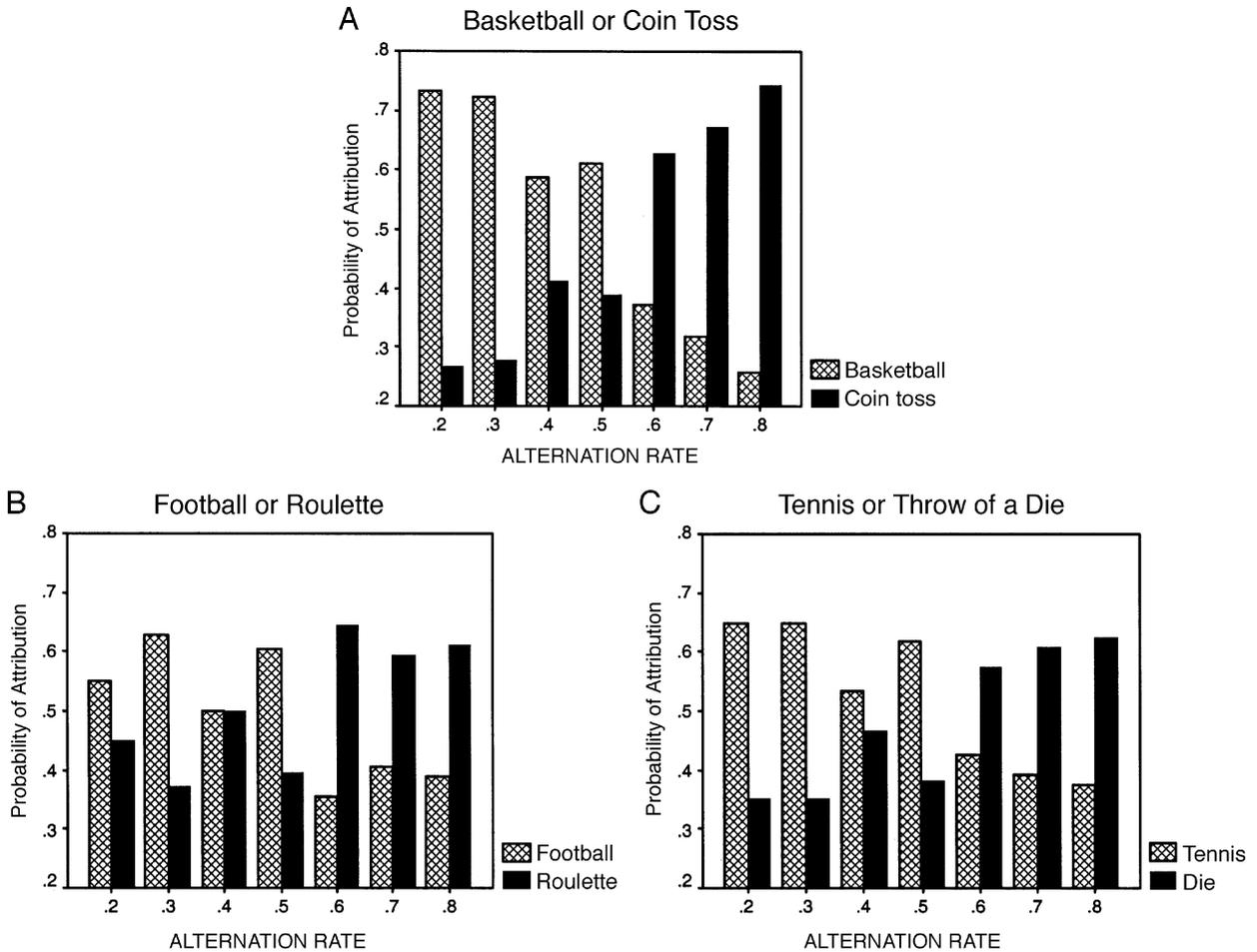### B     Football or Roulette

### C     Tennis or Throw of a Die

**Figure 5. (A) Probability of attributing sequences to basketball player scoring or coin toss. (B) Probability of attributing sequences to football team scoring or roulette. (C) Probability of attributing sequences to tennis player serving or throw of a die.**

less, plainly the two sequences are psychologically perceived quite differently; subjects *simultaneously* exhibited both positive and negative recency—the hot hand fallacy and the gambler's fallacy—for two binary sequences with identical statistical properties.

This result highlights the incompleteness of the representativeness account of misperceived randomness. As both effects can occur in expectations for *statistically* indistinguishable random sequences—and at the same time in the same head—some means is required to determine which expectation occurs when. Our alternative to the representativeness hypothesis—that both effects arise because people refer to a biased concept of randomness—is that people have different expectations for sequences based on their experiences with different kinds of events. More specifically, we propose that sequences of outcomes reflecting human performance yield anticipations of positive recency, whereas outcomes due to inanimate chance mechanisms yield anticipations of negative recency. The results of Experiment 1 are consistent with this notion but are limited in that we only investigated expectations to one of each type of process. In our second experiment, we examined expectations for a wider sample of sequence sources. The two tasks studied in Experiment 1 differ in response format and modality in that one task involves a binary choice, whereas the other involves a quantitative confidence estimate. Moreover, one probability is computed frequentistically (as a proportion of predictions across trials), whereas the other probability is based on confidence estimates within singular trials. In our second experiment, task demands and response assessment for different expectancies are matched.

### EXPERIMENT 2

In this experiment, we presented the subjects with three sets, each consisting of 28 binary sequences varying in recency. The subjects were told that each sequence showed output from one of two specified processes. A pair of processes—one representing human skilled performance and one representing chance performance—

was identified for each set of sequences: (1) a professional basketball player's scoring attempts during a game (hit and miss) or successive tosses of a coin (heads and tails); (2) a professional soccer team scoring in a series of matches (score and no-score) or successive outcomes of a roulette wheel (red and black); (3) a professional tennis player getting a first serve in (success and fail), or successive throws of a single die showing odd (1, 3, 5) and even (2, 4, 6) numbers. Their task was to identify which of the two named sources they thought was more likely to be responsible for each sequence. We hypothesized that sequences with positive recency would be attributed to human performance, whereas those with negative recency would be attributed to inanimate chance mechanisms.

## Method

**Subjects**. Thirty-three students from Ben Gurion University, Sapir and Achva colleges were recruited in return for credit points, as a partial completion of the requirements of the introductory psychology course.

**Design and Materials**. The 28 sequences in each set were computer-generated binary sequences. Each set comprised 4 exemplars of seven different alternation rates: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. To avoid any possible confounding between the symbols and the nature of the generating process (for instance, 1 could be intuitively related to *hit*, and 0 to *miss*), all events were "disguised" as @ and # symbols. Each series comprised 21 characters: 11 @ and 10 #. Thus the total number of each symbol across all series was kept constant.

There is a direct relation between the alternation rate of the 21 symbols and recency. If all 11 @ symbols were positioned in the first 11 places of the series, followed by the 10 # symbols, the series would have a single alternation. Of the 20 symbols followed by another symbol (all except the last symbol), 19 are followed by an identical symbol implying positive recency and the lowest alternation rate. If the 21 symbols were arranged so that none had a neighbor of the same type, this would result in negative recency and the highest alternation rate. To prepare the various series for the experiment, a computer program generated 21 random binary sequences (composed of 10 "0" and 11 "1" symbols). After generating each sequence, the software calculated the alternation rate and stored the series. The program stopped after finding 4 exemplars of each alternation rate.

Table 1 illustrates one exemplar of each of the 7 alternation rates in the disguised form utilized in the experiment (alternation rates were not identified or referred to). The same 28 sequences were presented in 3 different random orders for each set, and the order of presentation of the process pairs was also random across subjects.

**Procedure**. The subjects were invited to a regular lecture hall. An instruction sheet explained that the experiment was to test the human ability to recognize sequential patterns created by various sources. An example explained that "a coin tossed four times in a row may fall first on its Head, second on its Tail, next on its Tail again and finally on its Head. Other sequences may result from a basketball player either scoring or missing the basket, or from a roulette wheel's outcomes being either red or black."

The subjects were briefed that they would see binary sequences comprised of @ and # and that each sequence was a disguised output from one of two specified processes. Their task was to decide which of the two processes was most likely to account for the pattern observed in the outcome sequence. As added motivation, the subjects were promised an additional credit point if their number of correct responses was above average.[1] No feedback was given during the experiment.

The subjects were then given a three-page questionnaire, each page of which consisted of a short description of two processes, followed by the 28 sequences, each of which contained 21 symbols. Two empty response spaces next to each sequence were labeled to identify the two processes. The subjects were asked to examine each sequence individually and to use their intuition and best judgment to identify the source of the series by placing a mark in one of the labeled response cells. Completion of the task took about 20 min.

## Results and Discussion

We collated the data so as to be able to analyze how subjects' judgments of the source of sequences varied with alternations (see Figures 5A–5C). Data from 2 subjects were excluded from the analysis because both responded uniformly across all items. Missing data left us with usable responses from 29 subjects for the first two sets and 30 subjects for the third set.

We performed a one-way ANOVA on the decisions for each set of process pairs, treating the probability of attribution to chance process rather than human performance (subjects encountered each alternation rate four times within each set) as the dependent measure and the seven levels of alternation rate as the independent variable. These analyses confirmed a significant effect of alternation rate on probability of choosing skilled/chance performance for all three sets [basketball–coin, $F(6,168) = 14.26$, $MS_e = 0.09$, $p < .001$; football–roulette, $F(6,168) = 3.19$, $MS_e = 0.11$, $p < .005$; tennis–die, $F(6,174) = 4.39$, $MS_e = 0.11$, $p < .001$]. The linear trend across alternation rate was also significant for each pair [basketball–coin, $F(1,28) = 23.95$, $MS_e = 0.30$, $p < .001$; football–roulette $F(1,28) = 2.94$, $MS_e = 0.42$, $p < .05$, one-tailed test; tennis–die, $F(1,29) = 4.39$, $MS_e = 0.38$, $p < .02$].

These results reveal that, as hypothesized, for all three process pairs, subjects were more likely to attribute sequences with low rates of alternation (more "streaks") to

## Table 1
### Example of Sequences for Experiment 2

| Series | | | | | | | | | | | | | | | | | | | | | Alternation Rate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| @ | @ | @ | @ | @ | @ | @ | @ | @ | # | @ | # | # | # | # | # | # | # | # | @ | | 0.2 |
| @ | # | @ | # | # | # | # | # | # | # | @ | @ | @ | # | @ | # | @ | @ | @ | @ | @ | 0.3 |
| @ | @ | @ | @ | @ | # | # | # | # | # | @ | @ | # | @ | # | @ | # | @ | @ | # | @ | 0.4 |
| @ | # | # | @ | # | # | @ | # | # | # | @ | # | @ | @ | @ | # | @ | @ | @ | @ | @ | 0.5 |
| # | @ | # | @ | @ | # | @ | @ | # | @ | @ | # | # | @ | # | @ | @ | @ | # | # | # | 0.6 |
| # | @ | # | @ | @ | # | # | @ | # | @ | @ | # | @ | # | @ | @ | @ | @ | # | @ | # | 0.7 |
| @ | # | # | # | @ | @ | # | @ | # | @ | @ | # | @ | # | @ | # | @ | # | @ | # | @ | 0.8 |

human skilled performance, whereas sequences with high rates of alternation were attributed to inanimate chance processes. Evidently, subjects had disparate expectations for sequences produced by the two types of process we presented in the experiment. Moreover, in this experiment we used the same measure for the two forms of recency associated with different processes. The recency expectations are not symmetrical; sequences with zero recency (alternation rate of .5) were attributed to human skilled performance rather than inanimate chance. Studies of perceived randomness (e.g., Falk, 1975, 1981; Gilovich et al., 1985; Lopes & Oden, 1987) also show that alternation rates of .5 are not judged as random. The precise contribution of the differing expectations for each type of sequence to the judgments studied here remains to be determined; plainly though, and in keeping with our account of the gambler's fallacy and the hot hand fallacy, people have distinct prior expectations for the characteristic recency associated with different classes of processes generating sequential outputs.

## GENERAL DISCUSSION

The representativeness accounts for the gambler's fallacy and the hot hand fallacy propose that a single general—and erroneous—concept of randomness is the reference for both of these effects. According to this account, in both cases, subjects reject the random sequences they see as being unrepresentative of their faulty concept of statistical randomness. By believing that chance mechanisms should not exhibit long runs, the gambler's fallacy is invoked, whereas observing long runs of success refutes the notion that outcomes are random, and so the hot hand fallacy is invoked. However, without clarifying a mechanism to identify which of the two distinct and opposing prior expectations arises, there is an incomplete explanation of both the hot hand and gambler's fallacies with a single heuristic. The findings from our first experiment—simultaneous positive and negative recency for two statistically identical sequences—bring that difficulty into sharper focus. Without varying initial expectations about random sequences, the representativeness heuristic cannot explain why, for example, we do not fallaciously perceive runs of basketball scoring success as prognostic of failure, or roulette wheels as prone to getting hot.

Aside from its incompleteness, an explanation that, without specifying any conditions, predicts two opposite reactions to randomness prompts concerns about its testability. As Falk and Konold (1997) commented, "Although some testable implications of the representativeness heuristic have been repeatedly confirmed by Kahneman and Tversky [1972], and the concept was fruitful in inspiring a host of related studies, there is no established procedure for deducing how the heuristic will be implemented in a specific task"(p. 305).

A related question for the representativeness account is how people come to acquire their apparent misconception of randomness: Why do people believe that random sequences should exhibit local representativeness when in fact, of course, random sequences do not exhibit this characteristic? Although experiments show that people can learn about statistical randomness by observing it (e.g., Neuringer, 1989; Treisman & Faulkner, 1990), the representativeness heuristic explanation implies that this does not happen; real randomness does not reliably exhibit the local representativeness that the heuristic account assumes to underlie human expectations of random outcomes.

So what does prompt people's inappropriate responses to randomness? Our suggestion is that a biased concept of "pure" statistical randomness is not primarily responsible but separate (and somewhat valid) concepts of positive and negative recency that are cued when subjects decide which sort of previous experience the data are most likely to resemble. We speculate that the different apparent biases regarding statistical randomness may be acquired through life experience via sequences of events—negative recency in the natural ecology of uncertain events involving natural phenomena influenced by sampling without replacement (e.g., Ayton et al., 1989, 1991; Lopes, 1982) and the experience of positive recency in repetitions of human skilled performance with varying outcomes (Adams, 1995; Dorsey-Palmateer & Smith, 2004; Gilden & Wilson, 1995, 1996; Smith, 2003). Because it can hardly be reasonably denied that expectations are triggered by perceived similarity to previous experience, in some very general sense a representativeness explanation is inevitable (see Kahneman & Frederick, 2002, for discussion of a huge range of judgments where one attribute substitutes for, or represents, another). However, the experimental findings presented here show that, when presented with random sequences, people do not merely (or even) conclude that the sequences are not representative of statistical randomness. Instead, they decide what the data *are* representative of—namely, processes producing negative or positive recency.

Recently, Falk and Konold (1994, 1997) presented evidence that, when people are asked to explicitly judge the randomness of sequences, they use a measure of encoding the sequences' difficulty. They found that ratings of randomness were more strongly correlated with the difficulty of memorizing the sequences than with an entropy measure of randomness, leading them to conclude that subjects may base their judgments of randomness of a sequence on an implicit attempt to encode it. Sequences that alternate more than would be expected by chance are more difficult to memorize than those with chance alternations, which is consistent with people's judgments of randomness. These studies confirm that people have an erroneous concept of randomness, and suggest a basis for representativeness, but they do not address the issue of which competing hypotheses people may invoke when judging that a sequence is not random.

Other research relevant to a consideration of the difference between the hot hand and gambler's fallacies is reported by Wagenaar (1988, Chapter 7). Wagenaar noted

that in games of chance such as roulette, the outcomes of the wheel are typically held by gamblers to be random and governed by unpredictable chance. They may understand that the statistics of the wheel are against them, yet they bet because the outcome of a gamble is seen as the result of an interaction between two events: a player's choice of number and the outcome of a random device. According to this perspective, although analysis of the mathematics of the outcomes of the wheel may offer no hope, this says nothing about peoples' choices. Players apparently believe that their choices of number to bet on can be "lucky." Consequently, the outcomes of betting are commonly seen as governed by luck, which is not thought of in the same way as chance (see Wagenaar & Keren, 1988). Many gamblers see luck as something "streaky" that can be recognized and exploited. For example, if you notice it is your lucky day, you can take advantage of games of chance and win, or, if you notice that your luck is out, you should quit. In games such as roulette, such beliefs are, of course, nonnormative because runs of winning and losing show the same (zero) sequential dependency as the runs of outcomes on the wheel such as odd or even numbers. Others have noticed the parallel between the concept of luck and the hot hand (cf. Falk & Konold, 1997). Note that the distinction identified by Wagenaar and Keren is, like our account of the two fallacies studied here, partitioned between (random) outcomes for inanimate mechanisms and actions performed by people.

Other research is consistent with our empirical findings. Burns and Corpus (2004) have shown that respondents assume positive recency (akin to the hot hand) for forecasting scenarios they judged as nonrandom and negative recency (akin to the gambler's fallacy) for scenarios they rated as random. Boynton (2003) has also shown that success increases (and failure decreases) confidence in predicting random series, while explicit instructions that the source generated random output increased the likelihood of predicting alternations. Sundali, Croson, and Gold (2000) observed that casino roulette gamblers consistently bet more after winning than after losing, whereas the gambler's fallacy was less consistent: About half preferred to choose numbers that had not previously appeared, and half preferred numbers that had previously appeared.

A rather striking example of a similar dissociation reported by Perruchet (1985) was discussed recently by Clark, Manns, and Squire (2001). These authors experimented with an eyeblink conditioning paradigm. Human subjects were presented with a tone and, on 50% of trials, an airpuff in one eye, which eventually resulted in a conditioned eyeblink response to the tone. Whereas subjects' rated expectancy that a tone would be accompanied by an airpuff resembled the gambler's fallacy (decreasing expectancy after a run of airpuffs, increasing expectancy after a run of no-airpuff trials) their frequency of eyeblink responses showed the opposite pattern. Thus, when the likelihood of a conditioned response increased, the expectancy decreased and vice versa. Although the

cause of this dissociation is not understood, it is clear that the anticipatory response to random sequences assessed by two different means can simultaneously oppose each other.

Finally, let us make clear that we are not disputing the existence of the hot hand and gambler's fallacies—only their interpretation. The empirical case for the gambler's fallacy is strong and is not restricted to laboratory demonstrations; for example, Clotfelter and Cook (1993) revealed the effect in analyses of purchases of state lottery numbers. Although Gilovich et al.'s (1985) claim that there are no streaks in basketball has been greeted with skepticism (e.g., Hooke, 1989; Wardrop, 1995, 1998) and even disbelief (Larkey, Smith, & Kadane 1989), the statistical case has been robustly defended (Tversky & Gilovich, 1989b). Camerer (1989) also reported field evidence that betting on basketball games reflects a mistaken belief in winning streaks by teams, and Ayton (1998) cited evidence of similar misconceptions regarding soccer players' scoring performances.

Nonetheless, the consequences of such fallacious beliefs may not be so obvious as is often assumed. Burns (2003, 2004) claims that the hot hand fallacy permits successful "fast and frugal" judgments of the shooting percentage of individual basketball players but does not deny that belief in the hot hand is a fallacy. This distinction raises the deeper question of whether behavior should be judged on the validity of the beliefs on which it is based or on the adaptiveness of the resulting behavior. As allocation decisions (who should take the next shot?) based on observations of streaks would be better than those made by ignoring this cue, an invalid belief can lead to adaptive behavior. Studying beliefs by comparison with normative models and studying the adaptiveness of their associated behaviors can lead to dissociable conclusions about the efficacy of cognition.

English cricketer David Gower was once asked about his chances of scoring at least 100 runs in his next test-match, as he had in his previous two. He replied that the odds of repeating his success must be lengthening but then, on the other hand, because he was "on a roll," perhaps the odds should be shortened. His response illustrates the opposing expectations that are cued depending on how the event is categorized. Assuming that cricketers' scores in successive innings are independent (cf. Burrows & Talbot, 1985), Gower appears to have been successively entertaining the two opposing biased expectations that we call the gambler's fallacy and the hot hand fallacy. Echoing Laplace, we suggest that it is principally in situations involving conditionally independent events that two different randomness illusions can deceive.

## REFERENCES

ADAMS, R. M. (1995). Momentum in the performance of professional tournament pocket billiards players. *International Journal of Sport Psychology*, **26**, 580-587.

AYTON, P. (1998, September 19). Fallacy football. *New Scientist*, **159**, 52.

AYTON, P., HUNT, A. J., & WRIGHT, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, **2**, 221-238.

AYTON, P., HUNT, A. J., & WRIGHT, G. (1991). Randomness and reality. *Journal of Behavioral Decision Making*, **4**, 222-226.

BAR-HILLEL, M., & WAGENAAR, W.-A. (1991). The perception of randomness. *Advances in Applied Mathematics*, **12**, 428-454.

BOYNTON D. M. (2003). Superstitious responding and frequency matching in the positive bias and gambler's fallacy effects. *Organizational Behavior & Human Decision Processes*, **91**, 119-127.

BUDESCU, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 25-39.

BURNS, B. D. (2003). When it is adaptive to follow streaks: Variability and stocks. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

BURNS, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand." *Cognitive Psychology*, **48**, 295-331.

BURNS, B. D., & CORPUS, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand." *Psychonomic Bulletin & Review*, **11**, 179-184.

BURROWS, B. L., & TALBOT, R. F. (1985). Boycott, Botham and the exponential distribution. *Teaching Statistics*, **7**, 42-47.

CAMERER, C. F. (1989). Does the basketball market believe in the "hot hand." *American Economic Review*, **79**, 1257-1261.

CLARK, R. E., MANNS, J. R., & SQUIRE, L. R. (2001). Trace and delay eyeblink conditioning: Contrasting phenomena of declarative and nondeclarative memory. *Psychological Science*, **12**, 304-308.

CLOTFELTER, C. T., & COOK, P. J. (1993). The "gambler's fallacy" in lottery play. *Management Science*, **39**, 1521-1525.

DERKS, P. L. (1962). The generality of the "conditioning axiom" in human binary prediction. *Journal of Experimental Psychology*, **63**, 538-545.

DERKS, P. L. (1963). Effect of run length on the gambler's fallacy. *Journal of Experimental Psychology*, **65**, 213-214.

DORSEY-PALMATEER, R., & SMITH, G. (2004). Bowlers' hot hands. *The American Statistician*, **58**, 38-45.

EDWARDS, W. (1961). Probability learning in 1,000 trials. *Journal of Experimental Psychology*, **62**, 385-394.

ESTES, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89-128). New York: Academic Press.

FALK, R. (1975). Perception of randomness. Doctoral dissertation (in Hebrew with English abstract). Jerusalem: Hebrew University.

FALK, R. (1981). The perception of randomness. In *Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (pp. 222-229). Grenoble, France.

FALK, R., & KONOLD, C. (1994). Random means hard to digest. *Focus on Learning Problems in Mathematics*, **16**, 2-12.

FALK, R., & KONOLD, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, **104**, 301-318.

GILDEN, D. L., & WILSON, S. G. (1995). On the nature of streaks in signal-detection. *Cognitive Psychology*, **28**, 17-64.

GILDEN, D. L., & WILSON, S. G. (1996). Streaks in skilled performance. *Psychonomic Bulletin & Review*, **2**, 260-265.

GILOVICH, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

GILOVICH, T., VALLONE, R., & TVERSKY, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, **17**, 295-314.

HOOKE, R. (1989). Basketball, baseball, and the null hypothesis. *Chance: New Directions for Statistics & Computing*, **2**, 35-37.

JARVIK, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, **41**, 291-297.

KAHNEMAN, D., & FREDERICK, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.

KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-454.

LAPLACE, P. S. DE (1951). *A philosophical essay on probabilities*. New York: Dover. (Original work published 1796)

LARKEY, P. D., SMITH, R. A., & KADANE, J. B. (1989). It's okay to believe in the "hot hand." *Chance: New Directions for Statistics and Computing*, **2**, 22-30.

LEE, W. (1971). *Decision theory and human behavior*. New York: Wiley.

LOPES, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 626-636.

LOPES, L. L., & ODEN, G. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 392-400.

NEURINGER, A. (1989). Can people behave randomly? The role of feedback. *Journal of Experimental Psychology: General*, **115**, 62-75.

PERRUCHET, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Sciences*, **20**, 163-170.

PINKER, S. (1997). *How the mind works*. New York: Norton.

SMITH, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin & Review*, **10**, 753-758.

SUNDALI, J., CROSON, R., & GOLD, E. (2000, May) The gambler's fallacy versus the hot hand: Empirical data from casinos. *Seventh Conference on Behavioral Decision Research in Management*. Tucson, Arizona.

TREISMAN, M., & FAULKNER, A. (1990). Generation of random sequences by human subjects: Cognitive operations or psychophysical process? *Journal of Experimental Psychology: General*, **116**, 337-355.

TVERSKY, A., & GILOVICH, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance: New Directions for Statistics & Computing*, **2**, 16-21.

TVERSKY, A., & GILOVICH, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance: New Directions for Statistics & Computing*, **2**, 31-34.

WAGENAAR, W.-A. (1988). *Paradoxes of gambling behavior*. Hove, U.K.: Erlbaum.

WAGENAAR, W.-A., & KEREN, G. (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making*, **1**, 65-75.

WARDROP, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, **49**, 24-28.

WARDROP, R. L. (1998). Basketball. In J. Bennett (Ed.), *Statistics in sport* (pp. 65-82). London: Arnold.

WITTE, R. S. (1964). Long-term effects of patterned reward schedules. *Journal of Experimental Psychology*, **68**, 588-594.

**NOTE**

1. Because the sequences were artificially generated, there were no correct answers, so we scored responses according to our hypothesis about how recency would be associated with the sequences.