

What's Next? Judging Sequences of Binary Events

An T. Oskarsson, Leaf Van Boven,
and Gary H. McClelland
University of Colorado at Boulder

Reid Hastie
University of Chicago

The authors review research on judgments of random and nonrandom sequences involving binary events with a focus on studies documenting gambler's fallacy and hot hand beliefs. The domains of judgment include random devices, births, lotteries, sports performances, stock prices, and others. After discussing existing theories of sequence judgments, the authors conclude that in many everyday settings people have naive complex models of the mechanisms they believe generate observed events, and they rely on these models for explanations, predictions, and other inferences about event sequences. The authors next introduce an explanation-based, mental models framework for describing people's beliefs about binary sequences, based on 4 perceived characteristics of the sequence generator: randomness, intentionality, control, and goal complexity. Furthermore, they propose a Markov process framework as a useful theoretical notation for the description of mental models and for the analysis of actual event sequences.

Keywords: hot hand, streaks, gambler's fallacy, binary sequence, Markov process

In mortar attacks . . . lie down. If you can, crawl into one of the holes made by previous shells because . . . lightning rarely strikes twice in the same place. (Fisher, 1999, p. 36)

When people observe events in the world, they often perceive them as binary sequences occurring over time. For example, births in single families are often seen as a sequence of boys and girls, a 7-day weather forecast is expressed as a sequence of rainy or sunny days, the stock market's performance can be summarized as a series of ups and downs, and a basketball player's shots in a game can be perceived as a sequence of hits and misses while the game outcomes are recorded as a series of wins and losses.

Thinking about the relationships between events and the structures of sequences is a central capacity underlying human adaptive behavior. People's beliefs about future events affect important decisions and behaviors. For example, are parents less likely to have another child, if they predict it will be yet another boy? Will the hurricane hit my house this time? Is it the right time to sell my stocks? Will the Cubs blow it again?

The capacity to make these judgments is partly automatic. Two-month-old infants make anticipatory eye movements after a few minutes of exposure to alternating visual stimuli (Canfield & Haith, 1991), and adults' hemodynamic responses in the right

prefrontal cortex positively correlate with violations of anticipated sequence patterns (Huettel, Mack, & McCarthy, 2002). Nevertheless, most of the research on judgments of what's next in binary sequences engages deliberate inference and judgment processes. Studies examining perception of two types of sequences have dominated behavioral research. First, there is a long history of studies of judgments about events produced by random mechanisms such as coin tosses, roulette wheels, and biological birth processes (Ayton, Hunt, & Wright, 1989; Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Nickerson, 2002).¹ Second, following Gilovich, Vallone, and Tversky's (1985) seminal study of basketball shooting, there are many studies of sequences occurring in popular sports events and people's perceptions of those events (Alter & Oppenheimer, 2006; Bar-Eli, Avugos, & Raab, 2006).

In studies of both random mechanisms and skilled sports performances, people's judgments and their underlying beliefs depart systematically from the actual patterns observed in both types of sequences. Furthermore, people's expectations differ dramatically for random sequences versus sports performance sequences. When evaluating sequences generated by a random mechanism, people believe that streaks of events will be shorter than they are in true Bernoulli binomial sequences; this judgment phenomenon is sometimes referred to as the *gambler's fallacy* (Tune, 1964). Thus, a person tends to believe that the chance of getting a tail on a coin toss increases after three heads have appeared on the previous tosses, despite the fact that each toss is independent of the other tosses and the probability of a head is constant at .50, regardless of previous outcomes. In contrast, in the sports domain people exhibit a *hot hand fallacy*, expecting streaks of successes in

An T. Oskarsson, Leaf Van Boven, and Gary H. McClelland, Department of Psychology, University of Colorado at Boulder; Reid Hastie, Chicago Booth Graduate School of Business, University of Chicago.

Portions of this article were based on An T. Oskarsson's doctoral dissertation and were presented at the annual meeting of the Society for Judgment and Decision Making, Toronto, Ontario, Canada, November 2005. We are grateful to Eugene Caruso, Nick Epley, Tom Gilovich, Denis Hilton, Carol Krumhansl, and Jack Soll among many others who have provided useful feedback on some of the ideas presented in this article.

Correspondence concerning this article should be addressed to An T. Oskarsson, who is now at the Department of Psychology, University of California, Davis, One Shields Avenue, Davis, CA 95616. E-mail: an.oskarsson@colorado.edu

¹ The philosophical issues raised by the concept of undetermined, random events are outside the scope of this review (see Bennett, 1999, for an introduction to the concept of randomness). It is sufficient for our purposes to associate these events with the behavior of random generating devices and to label a novel sequence as *random* if it is consistent with the behavior of a Bernoulli process (see *Models of Random Processes* section below).

performance to be longer than they are in fact (Gilovich et al., 1985). For example, people believe that a basketball player's chances of making a shot are higher if the player has just made the previous three shots than if he had missed or had mixed success on the previous three shots. But, surprisingly, the preceding shots seem to be unrelated to the next outcome, and basketball shooting sequences are statistically similar to coin tossing sequences.

Before we move on, we should clarify the referents of the terms we use throughout this article. The term *gambler's fallacy* implies that the judgments are costly to the observer; but like many other researchers, we use both the terms *gambler's fallacy* and *negative recency* to refer to expectations that a streak of events will end (and we specify whether such judgments are costly or irrational). We will use the terms *streak* or *positive recency* to refer to judgments that a streak of events will continue. Note that the closely related term *hot hand* refers to a streak of only successful outcomes for one actor—a hot player making points in an athletic contest, a hot gambler winning in a casino game. We restrict our use of the term *hot hand* to this specific type of streak. Finally, we will make clear whether our references to sequences refer to the actual events in the external world or to beliefs or judgments about those events by an observer. For example, the term *hot hand* refers to a streak of successes in actual performance, while *hot hand belief* refers to an expectation that a streak of successes will continue.

This article is organized as follows. We begin with a review of the empirical findings from the past 50 years of behavioral research on judgments of mechanically produced binary sequences, sports events, and financial markets. For each domain, we summarize findings concerning the structural properties of actual event sequences and then what is known about people's judgments of those sequences. In the second section, we review theoretical proposals to account for people's beliefs about sequences and then introduce an original framework to account for people's judgments in terms of explanations based on beliefs about the generating mechanisms.

Altogether our review of research on sequence judgments led us to make three recommendations. First, more research is needed on judgments of sequences outside of the two overstudied domains of random mechanisms and skilled sports performances. Second, we favor an account of sequence judgments in terms of the mental models people learn and create to explain how the events are generated in each type of sequence. These naive explanations seem to depend largely on a set of basic process characteristics: random/nonrandom, intentional/unintentional, and controlled/uncontrolled. Third, researchers need to converge on a common theoretical framework to describe these mental model explanations, and we recommend the graphical displays and vocabulary used by probability theorists and engineers to describe Markov processes.

Review of Behavioral Studies

Random Sequences

In 2005, the number 53 had not been drawn in the Italian national lottery for almost 2 years. People began believing that there was an increased chance of the number 53 being drawn and as a result many Italians bet their life savings on 53 (Arie, 2005; BBC News, 2005). One woman drowned herself, leaving behind a

note saying she had secretly spent the entire family savings on her "number 53 habit." In the weeks before 53 was finally drawn, frustration with debts incurred from repeatedly betting on the number led one man to be arrested for beating his wife, while another man shot his wife and son before killing himself. In what became a national obsession, the country as a whole spent an estimated €3.5 billion on the number 53.

Models of Random Processes

Mathematicians and philosophers have been interested in people's subjective conceptions of randomness, almost since the inception of formal theories of random processes (Bennett, 1999). The philosopher/mathematician Pierre-Simon Laplace (1825/1995) provided a thoughtful, thoroughly modern discussion of the gambler's fallacy in his essay on "illusions in the estimation of probabilities," citing examples from gambling games, public lotteries, and births. More recently, Hans Reichenbach (1934/1949) noted that people expect a sequence of random Bernoulli trials, like coin toss or roulette wheel outcomes, to alternate more frequently than they do. Since the 1950s, there has been a thread of behavioral analyses of judgments of subjective randomness.

We use the term *random* to refer to the sequence of events generated by a mechanical or biological device that are causally and statistically independent from one another and that can be modeled as a series of Bernoulli independent and identically distributed trials² (Feller, 1950). In probability theory, a sequence or other collection of random variables is independent and identically distributed (i.i.d.) if each event (or variable) has the same probability distribution as the others and all are mutually independent. Many useful statistics describing the structure of a sequence of Bernoulli trials can be derived from elementary laws of probability theory, which we can compare to statistics summarizing event outcomes in an empirical sequence to see whether the sequence behaves randomly (see also Albert & Williamson, 2001). One such statistic is the probability of alternation, $p(A) = (r - 1)/(n - 1)$, for sequences of length n and with the number of runs (i.e., streaks or unbroken subsequences of a single outcome) equal to r . For example, suppose we tossed a coin 11 times and observed the following sequence of heads (h) and tails (t): hhhthtt-thht. For this sequence, $n = 11$, $r = 6$; therefore, $p(A) = (6 - 1) / (11 - 1) = .5$. Thus, this sequence has a $p(A)$ consistent with one produced by a Bernoulli i.i.d. process, because we know that the $p(A)$ of a Bernoulli sequence should be close to .5 when the binary outcomes are equiprobable.³ There are other se-

² We do not mean to say that random events are outside the realm of physical space/time/causality as we understand it from modern physics. In fact, careful analysis of any known mechanism that is used to produce random events demonstrates its underlying causal structure (e.g., Bayer & Diaconis, 1992; Lopes, 1982; Marsaglia, Zaman, & Tsang, 1990). Thus, when we use the term *random event* we mean an event that is described by idealized probability theory models such as those labeled *Bernoulli processes* (cf. Poincaré, 1914/1952).

³ Note that the value of $p(A)$ refers to the actual degree of alternation of some given sequence, or the proportion of alternations of a sequence, rather than the probability of alternations, which implies uncertainty. However, we will continue to use the term $p(A)$ to be consistent with terminology used by previous researchers.

quence statistics besides $p(A)$, such as the base rate, average run length, or average number of runs; however, all of these statistics are highly correlated.

Judgments of Random Sequences

Sequences produced by ideal random mechanisms. People have been asked to generate random sequences (Neuringer, 1986; Rapoport & Budescu, 1992, 1997; Treisman & Faulkner, 1990; Wagenaar, 1972), to detect whether a given sequence was produced by a random or nonrandom process (Falk, 1975, 1981; Lopes & Oden, 1987), to rate sequences for their perceived randomness (Falk, 1975), to predict future outcomes given some sequence (Ayton & Fischer, 2004; Gold, 1997; Gronchi & Sloman, 2008; Matthews & Sanders, 1984; Tyszka, Zielonka, Dacey, & Sawicki, 2008), and to assign subjective probabilities to future outcomes (Kahneman & Tversky, 1972; McClelland & Hackenberg, 1978). Participants were usually instructed to imagine a series of fair coin tosses to illustrate the concept of randomness, although other random mechanisms such as card decks, balls drawn from urns, throws of a die, roulette wheels, birth sex, and computer-generated stimuli have also been cited in instructions (for reviews, see Ayton et al., 1989; Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Nickerson, 2002). A slight task variation on generation of random sequences was studied by O'Neill (1987), Rapoport and Budescu (1992), and Budescu and Rapoport (1994). Participants were observed playing versions of a "matching pennies" game in which players have an incentive to make unpredictable choices. In trying to produce an unpredictable series of choices, participants exhibited a weak tendency toward negative recency, and alternations were slightly (but reliably) higher than .50.

The general conclusion from several dozen behavioral studies is that people do not have a statistically correct concept of random i.i.d. sequences (e.g., Ayton et al., 1989; Bar-Hillel & Wagenaar, 1991; Budescu, 1987; Falk & Konold, 1997; Lopes & Oden, 1987). For example, a study in which participants verbally reported their beliefs while generating random sequences of heads and tails found that large numbers of participants believed that the outcomes were dependent on one another (Ladouceur, Paquet, & Dubé, 1996). One consistent finding is that people associate randomness with negative recency, and they expect the outcomes in a random sequence to contain shorter streaks and alternate more than they should if the sequence has been produced by a Bernoulli process—confirming Laplace's (1825/1995) description of the gambler's fallacy. When people are shown proper Bernoulli sequences, they deem them to be nonrandom because the sequences seem to have streaks that are too long.

Some researchers argued that perception tasks are more appropriate than generation tasks for examining people's intuitions about randomness (Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997). The argument, by analogy, is that one does not need to be able to reproduce a scene or painting to be able to perceive or judge it, and so one does not need to be able to produce a random sequence to judge the randomness of sequences accurately. However, evidence for the association between subjective randomness and negative recency is also

found in perception tasks as well as generation tasks (Wagenaar, 1970, 1972).

In their review of subjective randomness research, Falk and Konold (1997) interpreted the findings of several studies in terms of the probability of alternation ($p[A]$) underlying participants' responses when performing various perception and generation tasks. In almost all the studies reviewed, participants perceived sequences with $p(A) = .6$ as most random and generated random sequences with $p(A) = .6$ or greater (e.g., Bar-Hillel & Wagenaar, 1991; Rapoport & Budescu, 1992, 1997). Thus, it appears that people have a miscalibrated conception of randomness such that binary sequences with $p(A) = .5$ are considered less random than sequences with $p(A) = .6$ (which in actuality has too many alternations and too short streaks to be random). Kareev (1992) has argued that the essential problem is that most researchers have requested short sequences, and under this condition, the participant correctly produces a typical sequence (e.g., generating five heads for a sequence of 10 coin tosses, resulting in too many alternations on average). However, people produce too many alternations even when generating very long sequences under the "be random" instruction.

In a perception experiment conducted by Lopes and Oden (1987), participants were asked to classify computer-generated sequences as having been produced by either a random process or by a nonrandom process. For half the participants, the nonrandom generator produced sequences with a higher degree of alternations than the random generator; for the other half, the nonrandom generator produced sequences with a higher degree of streakiness. The researchers found that the participants classified sequences with many alternations as random (unless the alternations were excessive) and associated sequences with long streaks and symmetrical patterns (e.g., cyclic or mirror patterns) with nonrandomness (see also Rapoport & Budescu, 1997). Also, participants had more difficulty discriminating between sequences produced by the random versus alternating generator than with sequences produced by the random versus streaky generator, although the two nonrandom processes equally deviated from random. The notion of miscalibrated randomness can explain this result: If participants were expecting $p(A) = .6$ for randomly generated sequences, then sequences produced by the alternating process should be more confusable with the truly random sequences than the sequences produced by the streaky process. Their results also suggest that people are more sensitive to long streak sequences than to frequent alternation patterns.

Other research has directly examined whether streaks are more salient in sequence judgments than patterns of alternations. Falk and Konold (1997) asked participants to memorize and later copy down binary sequences from memory, as well as to rate the sequences for apparent randomness. They found that the higher the number of alternations in the sequences, the longer participants needed to view the sequence prior to copying it down. Falk and Konold argued that streaks were more easily encoded and recalled as compared to alternations, and that streaks of the same outcome were readily utilized in judgments of nonrandomness, whereas streaks had to be relatively lengthy before evoking judgments of nonrandomness. Olivola and Oppenheimer (2008) found that memory for sequences of

random events showed biases consistent with the negative recency/gambler's fallacy expectation. The lengths of streaks present in the original sequence were underestimated—and when a streak was present early or late in a 25-event sequence the overall sequence was judged as less likely to be random, compared to when the same streak occurred in the middle of the sequence.

The salience of streaks in behavioral findings is consistent with the results of a functional magnetic resonance imaging study conducted by Huettel et al. (2002). Because research has indicated that the prefrontal cortex plays a central role in processing stimulus context and dynamic prediction, participants' hemodynamic response in their right prefrontal cortex was measured as they viewed random sequences of circles and squares. Greater activation was found when participants observed the end of a streak sequence (e.g., XXXXXXO) than the end of an alternating sequence of equal length (e.g., XOXOXOO). In fact, a significant increase in activation occurred after a streak of only two identical events (i.e., XXO), while the end of an alternating sequence of at least six events (i.e., XOXOXOO) was needed to elicit a significant increase in blood oxygen. The researchers concluded that the prefrontal cortex makes predictions about outcomes based on the pattern of events preceding an event and that increased activation in the prefrontal region indicates moment-to-moment updating of models of event patterns.

While most of the subjective randomness tasks required bottom-up, sequence-driven judgments (e.g., "Is this sequence random?"), participants have also been asked to perform tasks involving top-down judgments where they predict the next event in a sequence with some information about how it was generated. Studies of this type were popular in the late 1950s and 1960s in the probability learning paradigm used to test mathematical learning theory models (see Estes, 1976, for a review). In these tasks participants usually predicted which one of two novel events (e.g., left/right light, $\sqrt{}/+$ symbol) was about to occur and then received outcome feedback. In contrast to the studies just reviewed in which random generation mechanisms were explicitly specified, participants were usually given vague instructions about how the sequences were generated (e.g., "There is no pattern or system you can use that would make it possible to get all of your answers correct"; Edwards, 1961, p. 386) and were told "to get as many predictions correct as possible." The focus of these experiments was on a response pattern labeled *probability matching* in which, over many trials, participants' responses are allocated in proportions that match the outcome occurrence rates. For example, in a signal detection study by Healy and Kubovy (1981) that had monetary payoffs for judgment accuracy, participants exhibited almost perfect probability matching across a range of base rate outcome proportions. Probability matching is of considerable theoretical interest because it provides an apparent violation of rational payoff maximizing behavior in a simple "predict what's next" situation in which there is an extended opportunity for the participants to learn the optimal strategy. But, the story is much more complex when the full range of empirical studies is considered.

First, it turns out that although probability matching is a common finding, it is far from universal, and the exact recipe to

produce it is still unknown (Shanks, Tunney, & McCarthy, 2002, is the latest word in the search for necessary conditions for matching and for maximizing). Second, there are interpretations of the experimental task that can rationalize even consistent probability matching results, such as those reported by Healy and Kubovy (1981). Excellent reviews of the current state of this research are available elsewhere (Shanks et al., 2002; Vulkan, 2000), so we will only mention a few conclusions most relevant to the present review. As noted above, in many studies the nature of the outcome generating device was specified only vaguely and this seems to have encouraged participants to overthink to attempt to learn patterns (that were not present) that could support high levels of performance (cf. Feldman, 1959; Morse & Rundquist, 1960; Restle, 1961; Vulkan, 2000). Much of what appears to be mindless probability matching ignores the detailed structure of participants' responses and their reports of actively testing hypotheses in a fruitless search for patterns. Furthermore, even learning the simple base rates of different outcomes is difficult under outcome feedback conditions, although human participants still look like slow learners in these tasks. Note that learning involves a mixture of two motives: exploring the sequence to induce patterns and exploiting the sequence by applying the currently most effective strategy. We suspect that under these conditions participants were engaged in different mixtures of learning strategies. Some were testing hypotheses in a quest for the complex sequential pattern, others were responding in a reinforcement conditioning or memory-for-salient sequences mode (cf. Altmann & Burns, 2005), and still others were assuming the generator was a familiar random mechanism and were reasoning according to their (erroneous) beliefs about how random devices behave.

Many of these studies find a negative recency/gambler's fallacy-type pattern in predictions (e.g., Anderson, 1966; Anderson & Whalen, 1960; Jarvik, 1951; Nicks, 1959), but many also find positive recency biases or both (Altmann & Burns, 2005; Derks, 1963; Edwards, 1961; Feldman, 1959; Lindman & Edwards, 1961). In many of these experiments, there were sequential dependencies in the actual outcome sequence, so it is difficult to label any of the responses as fallacies (cf. Fiorina, 1971; Vulkan, 2000). In cases where the generating device is unequivocally random, predictions are likely to show the negative recency expectation we have noted for other random processes.

A study by Boynton (2003) demonstrates the influence of beliefs about the specific generating mechanism. Participants were presented with a random binary sequence and were asked to predict the next outcome. One group was given no specific information about the nature of the sequence, while another group was told that the task was "like trying to guess heads or tails" on tosses of a fair coin. A third group was told that the sequence was generated by a fellow student. Each group performed 100 trials. Although all participants viewed randomly generated Bernoulli sequences, the group instructed to imagine a fair coin showed negative recency, consistent with other research on misconceptions of randomness. The group that was told another student had generated the sequence fell showed less negative recency than the coin group and more negative recency than the no-instruction group. All of these findings are

consistent with the interpretation that participants rely heavily on preconceptions about the kinds of sequence that will be generated by different mechanisms.

When participants are asked to predict the next event produced by an unequivocally random mechanism, the results are more consistent. In the most extensive study of this kind, Gold (1997) conducted 18 experiments in which he presented participants with a series of actual coin tosses. Gold manipulated the number and the similarity of the coins tossed, added interruptions to the timing of events in the sequences, and varied the number of experimenters tossing the coin(s). (Gold also used poker chips drawn from an urn and found similar results.) Participants predicted the outcome of each coin toss, and the dependent measure of critical interest was whether participants exhibited the gambler's fallacy after having just witnessed four heads in a row. Generally speaking, gambler's fallacy occurred as long as the streak of four occurred for a single coin (regardless of the number of experimenters tossing the coin). If multiple coins were tossed to produce a single sequence, participants showed less gambler's fallacy the more dissimilar the coins (e.g., a penny and a Susan B. Anthony dollar coin) but showed more gambler's fallacy when coins were drawn from a single box. Also, when a 24-min pause interrupted the streak of four heads, the participants no longer displayed the gambler's fallacy.

Research by Roney and Trick (2003) confirms the importance of the observer's grouping of events. When the next event prediction was made for a sequence presented as a continuous group, gambler's fallacy habits were observed; when the sequence was interrupted and the to-be-predicted event was conceptualized as a member of a new group, the prior event pattern had no effect on the prediction. A similar principle may also explain the failure to observe trial-to-trial dependencies in psychophysics experiments where each judgment trial tends to be conceptualized as independent of its predecessors (Colle, Rose, & Taylor, 1974).

Ayton and Fischer (2004) demonstrated that while a person may expect negative recency for a random sequence, he or she can simultaneously expect positive recency for a sequence of wins and losses based on predictions of the same random sequence (see Croson & Sundali, 2005, for similar findings in the field). They asked participants to make 100 bets predicting the color outcome of a simplified roulette wheel. Participants also indicated their confidence in each of their predictions. Note that the roulette wheel's color sequence and the win/loss sequence of participants' judgments are statistically identical: Both consist of equiprobable, independent outcomes. Analysis of participants' responses following streaks of outcomes of Length 1–5 revealed that predictions of the machine's behavior (the wheel color) were consistent with gambler's fallacy (i.e., the longer a streak of color outcomes, the less likely participants were to predict the streak would continue). But, participants' confidence ratings in the predictions exhibited a hot hand belief—confidence increased following a run of successful predictions and decreased following a run of failed predictions. In other words, people's predictions for a sequence produced by a random device showed negative recency, but people's beliefs for a statistically identical sequence of successes and failures of their predictions exhibited positive recency.

Lotteries. The journalistic account of the recent Italian National Lottery (mentioned at the beginning of this section) strongly implicates gambler's fallacy thinking in predictions of winning numbers. More systematic studies of the Maryland State lottery confirmed this expectation. Clotfelter and Cook (1993) used a measure based on the payout amount as an index of number popularity (across several bet types). Each winning number on a straight bet paid off \$500 to the winner. Betting on previously winning numbers dropped after they were drawn to approximately 60% of their previous popularity and then gradually returned to their original ambient popularity after 3 months. (Note that the payout was a flat \$500 regardless of the number of bettors who select the winning number, so there was no financial penalty for following a gambler's fallacy strategy in the Maryland lottery.)

Terrell (1994) followed up on the Clotfelter and Cook (1993) with a study of the New Jersey pari-mutuel lottery. Under a pari-mutuel scheme, overbetting popular numbers (and underbetting unpopular numbers) incurs a financial loss because the payoff pot is split among all the bettors who pick the winning number. Terrell used player winnings as an index of over- and underbetting specific numbers (the state divided 52% of the money bet on a day's number evenly among those who picked the winning number—higher payoffs per winner indicated the number was underbet; lower payoffs suggested the number was overbet). Again, bettors exhibited the gambler's fallacy: 25% fewer players bet on a number that had won in the past week than on numbers that had not won for more than 8 weeks. The slight reduction in the number of bettors avoiding a recent winner in this study, compared to the bettors in Clotfelter and Cook's Maryland study, could be due to the pari-mutuel nature of the New Jersey lottery, if players are aware that betting on popular winning numbers results in a lower payoff. Nonetheless, lottery play, like the outputs of other mechanical devices that are conceptualized as random, yields consistent evidence for gambler's fallacy betting habits.

Births. A random sequence that is often encountered in everyday life is the series of boy and girl births in a family. Kahneman and Tversky (1972) asked participants the following:

All families of six children in a city were surveyed. In 72 families, the exact order of births of boys [B] and girls [G] was G B G B B G. What is your estimate of the number of families surveyed in which the exact order of births was B G B B B G?

Even though the two orders of births are about equally likely, participants believed that the second order was much less likely (median estimate = 30 families) because it did not reflect the expected base rate of boys and girls. Participants also thought that the order B B B G G G was significantly less likely than G B B G B G, indicating that the order of births was another aspect of the sequences under consideration when making their estimates.

McClelland and Hackenberg (1978) examined people's subjective probabilities for birth sex by asking college students and Philippine villagers the following question:

If a family already has B boys and G girls and they were going to have another child, do you think it more likely that they will have (a) a boy than a girl, (b) a girl than a boy, or (c) a boy as likely as a girl?

The B and G in this question were replaced by all possible combinations of the numbers 0, 1, 2, and 3. (Some of the college

participants also answered similar questions about coin tosses, and responses to both sets of questions were similar.) Despite the fact that birth dependencies are negligible (Ben-Porath & Welch, 1976; Rodgers & Doughty, 2001), the majority of participants in both samples predicted that a family's sex composition would balance out such that the less numerous sex would be the more likely sex of the next child. A whopping 78% of Philippine villagers exhibited this gambler's fallacy belief, as compared to 35% of University of Colorado students.

Individual differences. Although most people exhibit the gambler's fallacy when judging random sequences, there are individual variations in this tendency. The study by McClelland and Hackenberg (1978) of birth sequence judgments illustrates the substantial individual differences in such judgments. In addition to the differences between the Philippine sample and the Colorado sample, it was found that within the Colorado sample of 250 students, 35% displayed the gambler's fallacy, 34% correctly responded that either sex was equally likely, 20% exhibited positive recency (i.e., responses favoring the more numerous sex in the family), and 10% of participants had responses that were inconsistent and unclassifiable.

Individual differences have also been documented in tasks involving judgments of coins and other idealized random devices (Budescu, 1987; Falk, 1975; Gold, 1997; Keren & Wagenaar, 1985; Tyszka et al., 2008; Wagenaar, 1972). A minority of participants, usually between 5% and 10% of any sample, exhibit a tendency toward positive recency instead of negative recency. Wagenaar (1988) examined the bets made on the color outcomes of a roulette wheel and found that 60% of bets were consistent with negative recency, while 40% of bets were consistent with positive recency. Benhsain, Taillefer, and Ladouceur (2004) found that as much as 80% of verbal comments mentioned interoutcome dependencies in a study of occasional gamblers in a laboratory setting; however, they did not distinguish between references to the gambler's fallacy and other types of dependencies. Friedland (1998) claimed that gambling predictions depend on whether the gamblers are luck versus chance oriented personalities, and other researchers have argued that personality traits predict individual perception of patterns in random visual displays (Jakes & Helmsley, 1986).

Sundali and Croson (2006) found that only about one half of casino roulette players showed a gambler's fallacy betting pattern. They proposed that gamblers have a miscellaneous collection of beliefs that can be evoked by subtle contextual conditions. In addition to the gambler's fallacy and hot hand, they found evidence for roulette players' beliefs in hot outcome and stock of luck superstitions. The *hot outcome* refers to expectations of positive recency in a random sequence; for example, believing that a streak of red outcomes from a roulette wheel is going to continue because red is hot. The *stock of luck* refers to expectations that a player's winning streak will end, consistent with the belief that a player has a fixed stock of luck and that once it is spent, the probability of winning decreases. Sundali and Croson examined videotapes of 18 hr of roulette wheel play in a real casino and found gambling behavior consistent with belief in the hot hand (betting more after winning than losing), the gambler's fallacy (betting on a number that had not recently appeared as if it were due to come up), and hot outcome (betting on numbers that had previously appeared

as if they were hot) preferences. Less evidence was found for a stock of luck notion (that a gambler is due to lose after a winning streak), but perhaps the winning streaks in the 18 hr of observed roulette play were not long enough to evoke this belief or perhaps early success simply increased the players' levels of aspiration and persistence.

Keren and Lewis (1994) found indirect support for the notion of hot outcomes. Participants were asked to estimate how many observations of roulette spins would be necessary to identify a hot number on a roulette wheel. Participants displayed a strong and pervasive tendency to underestimate the number of observations needed. This suggests that when people observe streaks in a gambling context, they are quick to believe in hot numbers and expect positive recency instead of negative recency. Perhaps participants expected short subsequences of the biased roulette wheel to be locally representative of the biased base rate, thus underestimating the length of the sequence needed to conclude that the wheel was biased. In this way, people's miscalibrated notion of randomness promotes a tendency to see patterns in sequences, especially when they have some a priori expectation or theory about the causal mechanism generating the sequence (see later discussion of Gilovich et al.'s, 1985, explanation for hot hand biases in sports' judgments).

Summary and comments. People act as if they expect the outcomes in binary-event sequences produced by putatively random devices to revert quickly towards a 50-50 base rate. They seem to believe that such sequences should exhibit more alternations, shorter streaks, and fewer symmetries than the actual sequences produced by a random, stationary, independent-trials Bernoulli process. The modal subjective probability of alternation $p(A)$ is about .60 (Budescu, 1987; Falk & Konold, 1997). This result is often labeled the gambler's fallacy and it increases with increases in streak length, such that people are virtually certain that a run of 10 heads will be followed by a tail. Streak patterns are more salient and influence sequence judgments more than alternation patterns. Although the gambler's fallacy is the norm, substantial individual differences have been observed within and between domains. We interpret these judgments as reflecting a generally shared, flawed belief about the nature of a random Bernoulli process.

Nonrandom Sequences

In 1997, the Detroit Red Wings swept the Philadelphia Flyers four games to none to win the Stanley Cup hockey championship. Mike Vernon, the Red Wings' goalie, played every single game in the four rounds of playoffs and was voted the most valuable player. Interestingly, Vernon was Chris Osgood's backup goalie during the regular season—Osgood had a better winning percentage and goals against average. But the Red Wings' coach felt that Vernon was hot at the start of playoffs and went with his hot goalie for all of the playoff games. Following the Stanley Cup championship, the Red Wings kept Osgood and sent Vernon, the most valuable player, to the San Jose Sharks (D. G. Morrison & Schmittlein, 1998).

The Structure of Actual Sports Performance Sequences

Many sports fans, commentators, players, and coaches share a belief that a player can have the hot hand and be "on fire," "in

the zone,” “on a roll,” “playing his ‘A game,’” or “unstoppable.” Yet the controversial study conducted by Gilovich et al. (1985) reported analyses of field goal shooting statistics of professional basketball players playing for the Philadelphia 76ers, the New Jersey Nets, and the New York Knicks (as well as the free throw data from the Boston Celtics) and concluded that basketball players do not get the hot hand (the same results were also reported in Tversky & Gilovich, 1989a). Little evidence was found for dependencies between shot outcomes, unusually long streaks, an unusual number of streaks, or nonstationarity (nor was there evidence that hit rates varied systematically within or across games). In fact, it was slightly more likely that players’ shot sequences exhibited negative recency. Gilovich et al. confirmed the disparity between people’s perceptions and actual data with a controlled experiment in which varsity college basketball players made free throws while both the players and observers predicted the outcome of each attempt. Both players and observers believed that some players were hot while shooting free throws, but only 1 out of the 26 players actually showed reliable positive dependencies between shots and an unusual number of streaks, overall providing no evidence for shooting streaks.

Koehler and Conley’s (2003) analysis of performance in the NBA Long Distance Shootout Contest is one of the most convincing follow-ups to the original Gilovich et al. (1985) analyses. In this contest, professional players attempted 25 uncontested field goal shots in an interval of 60 s. These conditions would seem to maximize the chances for momentum, motivational, and other hot factors to produce streaks in performance. However, several sophisticated statistical tests failed to detect any nonrandom shooting patterns. Koehler and Conley also analyzed a subset of shot attempts that immediately followed announcers’ comments about the player being “hot,” “on fire,” and so forth. But again, no statistical evidence of streakiness or other dependencies was found.

Since the publication of Gilovich et al.’s (1985) study, a host of studies has been conducted seeking evidence for the hot hand in a variety of sports. (Bar-Eli et al., 2006, provided a comprehensive summary of the quest for evidence of streakiness in sports performance, and there is an excellent Web site dedicated to hot hand news: <http://thehotband.blogspot.com>). The conclusions for basketball have withstood several critical attacks (e.g., Adams, 1992; Cornelius, Silva, Conroy, & Petersen, 1997; Larkey, Smith, & Kadane, 1989; Shaw, Dzewaltsoki, & McElroy, 1992; Tversky & Gilovich, 1989b). Furthermore, contrary to observers’ intuitions, researchers have failed to document evidence for unusual streaks in baseball hitting (Albert & Bennett, 2001; Albright, 1993; Frohlich, 1994; Stern, 1997), baseball scoring (Lock, 2003), professional golf (Clark, 2003a, 2003b, 2004), volleyball scoring (Miller & Weinberg, 1991), or baseball and basketball game wins (Chang, 2003; Richardson, Adler, & Hanks, 1988; Vergin, 2000).

In contrast, there is evidence for streaky performance in golf putting (Gilden & Wilson, 1995), bowling (Dorsey-Palmateer & Smith, 2004; Frame, Hughson, & Leach, 2004), billiards (Adams, 1995), horseshoes (Smith, 2003), tennis (Klaassen & Magnus, 2001; Silva, Hardy, & Crace, 1988), darts (Gilden & Wilson, 1995), and hockey goalie performance (D. G. Morrison & Schmittlein, 1998). Although a conjecture, it appears that

nonreactive, turn-taking, uniform-trial individual sports are likelier to show nonrandom sequential dependencies in performance than more reactive and chaotic team sports events. When the conditions for each trial are more uniform (putting, billiards, bowling, horseshoes, darts), there is evidence for streaks; when there are many external factors affecting performance (basketball, baseball, football), there seem to be no statistically reliable streaks (for similar speculations, see Bar-Eli et al., 2006; Kaplan, 1990).

A fascinating collection of studies examine behavior by professional athletes in situations in which it is optimal to behave randomly to strategically outperform competitors. Such situations arise in tennis service and penalty kicks in soccer, where the actor selects between right or left placement with the goal of preventing the opponent from anticipating that placement. Note that in these situations, there is virtually perfect control over the target outcomes. Interestingly enough, at the professional level of play in both tennis and soccer, performance matches the ideal of an i.i.d. Bernoulli process (Chiappori, Levitt, & Grosseclouse, 2002; Palacios-Huerta, 2003; Walker & Wooders, 2001).

The hot hand conception of sports performances has two separable components that violate the Bernoulli trials description of an i.i.d. sequence: the notion that a player is hot, referring to the *global base rate* level of performance during a game or other extended perceptual unit (violating Bernoullian stationarity) and the notion of *interdependency* or *momentum*, being on a roll or “unstoppable,” referring to the conditional $p(A)$ relationship (violating independence). Neither type of hotness seems to be present in the performance of basketball and baseball players (the most analyzed sports), though there are still skeptics who argue that the analyses to date lack power to identify true streakiness if it is present (e.g., Hales, 1999; Miyoshi, 2000).

Most prior studies have focused on the independent trials component, but some analysts have tested for global variations in base rate performance level as a model for hotness and also found no evidence of nonstationarity in basketball and baseball (Albert, 1993; Albert & Bennett, 2001; Albert & Williamson, 2001; Forthofer, 1991; Frame et al., 2004; Wardrop, 1999). This distinction between independence versus stationarity is especially significant when one turns to studies of people’s intuitive beliefs and judgments about streakiness in sports performance. For example, Burns’s (2004) claimed that it is adaptive for a player to react to a run of hits by feeding the ball to the seemingly hot player, because the run is valid evidence that the player has a latent (or game-specific) higher base rate of scoring (in contrast to the Gilovich et al., 1985, emphasis on intershot dependencies). Nonetheless, to our knowledge, no one has reported a convincing data-based argument in support of nonstationarity in actual performance, though we cited several reports (above) of nonindependence in turn-based sports like tennis serving, golf putting, and so forth.

Judgments of Sports Performance Sequences

Most of the studies that examine belief in the hot hand involve prediction of human performance in sports or gambling situations. The research tasks usually require participants to make top-down, deductive judgments based on intuitions and prior expectations. A list of studies of judgments of nonrandom sequence generators is

Table 1
List of Studies Examining Judgments of Nonrandom Sequence Generators

Sequence generator	Reference	Task
Sports		
Basketball shooting	Gilovich et al. (1985)	Classify sequence
	Ayton & Fischer (2004)	Classify sequence
	Caruso & Epley (2008)	Predict outcome
	Burns (2004)	Rate generator, predict, likelihood
	Tyszka, Zielonka, Dacey, & Sawicki (2008)	Predict outcome
Basketball games	Gronchi & Sloman (2008)	Predict outcome
Baseball swings	Camerer (1989)	"Point spread" data
Machine mimics basketball, baseball	Caruso & Epley (2008)	Predict outcome
Football games	Matthews & Sanders (1984)	Likelihood ratings
Football scoring	Ayton & Fischer (2004)	Classify sequence
Tennis serves	Ayton & Fischer (2004)	Classify sequence
Non-sports		
Roulette prediction success	Ayton & Fischer (2004)	Confidence ratings
Sales competition	Croson & Sundali (2005)	Predict outcome
	Burns (2004)	Rate generator, predict, likelihood
Stock prices	Burns (2003)	Predict outcome
Student generated	Boynton (2003)	Predict outcome
Computer generated	Lopes & Oden (1987)	Classify sequence
	Falk (1981)	Classify sequence

presented in Table 1 (see Bar-Eli et al., 2006, for a similar tabulation).

The sports announcers in Koehler and Conley's (2003) study of the NBA Long Distance Shootout Contest are not the only people who strongly believe in the hot hand. When Gilovich et al. (1985) asked college basketball fans "Does a player have a better chance of making a shot after having just made his last two or three shots?," 91% of participants answered yes. Most participants also believed that a player was more likely to make the second free throw after making the first free throw than after missing the first free throw (68%). In addition, 81% of participants believed that it was important to pass the ball to a player who had just made several shots in a row.

Gilovich et al. (1985) then presented basketball fans with six shooting sequences of 11 hits and 10 misses, differing in probability of alternation $p(A)$ from .40 to .90. Participants were asked to classify each sequence as chance shooting, streak shooting, or alternation shooting. *Chance shooting* was defined as a sequence that looked just like a sequence of coin tosses, while *streak* and *alternation shooting* were defined as having clusters of hits and misses that were longer or shorter, respectively, than the clusters of heads and tails found in coin tossing. The majority of participants considered sequences with $p(A)$ s of .60, .70, and .80 to be the best examples of chance shooting (i.e., sequences with too many alternations and too few streaks to be true i.i.d. sequences), which is consistent with earlier findings from the subjective randomness research. The tendency to label a sequence as streak shooting decreased with increases in the probability of alternation of the sequence, but participants' responses indicated that they were miscalibrated:

The sequence with $p(A) = .50$ (i.e., the actual random i.i.d. sequence) was classified as an example of streak shooting by 62% of the fans.

Gronchi and Sloman (2008) presented participants with filmed sequences of coin tossing or basketball shooting and found the standard tendency to expect too many reversals (negative recency) for coins and too much streakiness (positive recency) for basketball. They emphasized the point that both the context sequence of recent outcomes and the domain (coins, basketball) interact to determine predictions. While there is a general tendency for positive or negative recency for particular domains, predictions also depended on the particular context sequence (e.g., hhhhtttt vs. htththt).

Burns and Corpus (2004) presented participants with three different scenarios designed to vary in degree of perceived randomness: (a) spins of a roulette wheel, (b) basketball free throws made by "your little sister," and (c) whether or not a salesperson has more sales than a coworker each week. For each scenario, participants were told that after 100 outcomes had occurred with a base rate of 50%, a streak of four outcomes in a row had just occurred. Participants predicted the next outcome, estimated the percentage chance that the next outcome would be a continuation of the streak, and rated each scenario for randomness of the outcomes.

The salesperson scenario was rated as least random, the basketball scenario next, followed by the roulette scenario. Randomness judgments correlated with participants' streak predictions ($r = .29$), such that scenarios rated as less random were associated with more predictions that the streak would continue. More specifically, 60% of participants predicted that the streak would continue for the salesperson scenario, followed by 46% for the basketball

scenario and 13% for the roulette scenario. Burns and Corpus (2004) also manipulated whether each scenario was presented in either past or future tense and found that the future versions of the nonrandom scenarios (i.e., the sales and basketball scenarios) led to more positive recency judgments (streakiness) than past versions. No such temporal orientation difference was observed for the random roulette scenario. Burns and Corpus concluded that the perceived randomness of the underlying mechanism believed to generate the events in a sequence is the most important factor affecting people's streak judgments.

On a similar theme, Tyszka et al. (2008) showed participants sequences of 180 events and asked for predictions of the next event after every 10 events. They told participants that the events represented the outcomes of coin tossing (head/tail), a fortune-teller's predictions (correct/incorrect), weather events (rainy/sunny), and basketball shooting (hit/miss). Participants rated the first two generators as random and the second two as deterministic. They found effects of the generator, with reversals more likely to be predicted for random processes (coin, fortune-teller) and streaks more likely for deterministic processes (weather, basketball). They also found consistent individual differences in the tendency to expect reversals or streaks, and responses to a questionnaire about everyday sequences were correlated with these habits.

Ayton and Fischer (2004) studied differences in judgments about the behavior of humans versus inanimate objects. Participants were presented with 3 sets of 28 binary sequences varying in probability of alternation $p(A)$ from .2 to .8. The task was to decide which of two mechanisms produced each sequence. A different pair of mechanisms was cited for each set of sequences: basketball shooting or coin tossing, football team scoring (score/no score) or roulette spins (red/black), tennis player's serves (success/fault) or throws of a die (even/odd). Participants were more likely to attribute streaky sequences with fewer alternations (with $p[A] = .5$ or less) to skilled humans, whereas sequences with more alternations (with $p[A] = .6$ or higher) were attributed to the random inanimate devices. Ayton and Fischer argued that a critical determinant of whether people expect positive or negative recency is whether they are considering actions performed by people or outcomes produced by an inanimate mechanism.

Caruso and Epley (2008) asked participants to predict the next outcome following a streak of three hits by either a professional basketball player (Paul Pierce) or by a machine designed to mimic the shots of a professional basketball player. Participants expected positive recency for the basketball player and for the machine but thought that the run generated by the actual player would be longer than the run generated by the machine.

Finally, Choi, Oppenheimer, and Monin (2003) conducted a study in which participants tossed coins and predicted the outcome of each toss. One group was told that they would win a lottery ticket if 12 heads appeared (making heads the goal), while a control group just tossed coins and predicted outcomes. Participants in the experimental group displayed the gambler's fallacy after experiencing a losing streak of tails, but they exhibited the hot hand following a winning streak of heads (by predicting that the next toss was more likely to be a heads and continue the streak). This could be interpreted as showing how an agent's intention (to obtain heads) can lead to expectations of either positive recency (continuation of winning streaks) or negative recency (end of losing streaks).

People also see momentum effects (streaks) at the team and game level in sports like basketball, baseball, and football. Matthews and Sanders (1984) presented participants with all 32 possible five-outcome binary sequences, labeled as the win/loss records of 32 football teams. For each specific win/loss sequence, participants predicted the likelihood of the football team winning its next (sixth) game on a scale of 1 (*very likely*) to 5 (*very unlikely*). A second group of participants was presented with the same sequences but was told that they represented the win/loss record of a person betting on coin tosses. A linear multiple regression was performed for each participant, with each of the five outcomes in the given sequences serving as predictors of the participant's judgment. For each participant, the beta weights for each predictor outcome could be either positive (reflecting positive recency) or negative (reflecting negative recency).

Beta weights tended to be positive for the football-game group and negative for the coin-toss group. Summation of the beta weights per participant revealed that all participants in the football-game group had a positive sum of beta weights, while 19 of 22 participants in the coin-toss group had a negative sum. Furthermore, while both groups placed more weight on the more recent outcomes, the football-game group relied more heavily on the earlier outcomes in the sequence for their predictions than did the coin-toss group. The researchers argued that the causally linked football game data were more highly integrated by participants than noncausally linked, or independent, coin toss data. In other words, participants' prior intuitions about football led them to interpret the football game sequences as more patterned and less random.

Beliefs in streaks in team performance can have negative financial consequences. After examining gambling bets made on NBA games during 1983–1986, Camerer (1989) found that the bets placed on teams with winning streaks tended to lose while the bets placed on teams with losing streaks tended to win, because of the inflated point spreads (see also Badarinathi & Kochman, 1994; Brown & Sauer, 1989). The point spread is the number of points the favored team is expected to win by, and the size of the point spread depends on the dollar amount bet on each team. For example, if a team has a point spread of +6, the team would have to win by 7 points to pay a bettor. The point spread data indicate that bettors acted as if they overestimated the likelihood that winning or losing streaks would continue and thereby pushed the point spreads too far apart, so that bets against streaking teams won more than 50% of the time. In other words, the gamblers (and the odds makers) exhibited a hot hand bias.

Although the norm in sports performance judgments seems to be the overprediction of runs and streaks (positive recency), negative recency has been observed in some sports judgments. For example, Metzger (1985) studied betting at parimutuel racetracks and found that after the favorite had won in several early races, long shots were overbet in later races and vice versa (and a run of long-shot winners would shift betting to favorites). It is difficult to interpret this intriguing aggregate correlational observation, but it seems most consistent with a stock of luck bias. When favorites have won their share of the day's races, it is time to bet on long shots. Terrell (1998) found similar patterns in betting at pari-mutuel greyhound races.

Stock Prices

Common stock prices are events that are notoriously overinterpreted in causal terms, with many investors rejecting the random walk model and making spurious attributions for stock price behavior (DiFonzo & Bordia, 1997, 2002; Morris, Sheldon, Ames, & Young, 2005). Many investment advisors rely on either momentum investment strategies ("Prices will streak; buy recent winners") or contrarian investment strategies ("Prices will reverse; buy recent losers"; Conrad & Kaul, 1998). Of course, one distinctive property of stock prices, compared to the other types of sequences we have considered, is that if investors hold a consistent theory for price movements and make investments based on that theory, the investments will impact the actual prices.

Momentum effects (positive recency) have been documented in actual stock prices, such that average stock returns are dependent on past performance under certain conditions. Jegadeesh and Titman (1993) found that companies with high stock returns in the last 3 months to a year tend to continue to outperform companies with low stock returns in the same time period, favoring a momentum interpretation. Reversal effects (negative recency) have been observed as well; De Bondt and Thaler (1985, 1987) found that prices tended to reverse over longer time intervals (greater than 12 months). The complete picture for price movements is complex and appears to depend on company size, industry, and specific time horizons. Since our focus is on binary event series, we do not provide a thorough review of stock prices (see Moskowitz & Grinblatt, 1999, for an instructive analysis).

There is ample evidence of both momentum (hot hand) and contrarian (gambler's fallacy) beliefs and habits in investor behavior. For example, the *disposition effect*—the tendency to sell stocks that have recently appreciated and to hold stocks that have declined—has been interpreted with reference to beliefs about negative recency in price trends (e.g., Odean, 1998; Offerman & Sonnemans, 2004; Shefrin & Statman, 1985). Rabin (2002) and others (e.g., Cheng, Pi, & Wort, 1999; Mullainathan, 2002; Sirri & Tufano, 1998) have interpreted the tendency of investors to overrely on advisors (and funds) who have demonstrated a few recent successes as due to belief in streaks of success (like the hot hand belief in sports). However, the presence of hot hand streaks in actual investment fund performance is highly controversial (Carhart, 1997; Elton, Gruber, & Blake, 1996; Hendricks, Patel, & Zeckhauser, 1993; Metrick, 1999).

We know of only one study that examines judgments of prices presented as binary events. Burns (2003) proposed that the inferred variability of a generating process is used as a heuristic for determining when to follow streaks, arguing that higher variability leads to more hot hand belief. Participants were presented with descriptions of two stocks, one from a large well-established company and one from a small recently established company. They were told either that the price of both stocks had increased (positive streak) or decreased (negative streak) in each of the last 6 months. Participants decided which of the two stocks was more likely to increase in the next month, was a better short-term (6-month) investment, was a better long-term (10-year) investment, and was likely to have a more stable price. For all three investment time periods, more participants opted to invest in the larger company

following a negative streak, indicating an expectation that the losing streak would reverse for the larger company but would continue for the smaller company. However, the pattern of responses was different when the companies were on winning streaks: Participants predicted that the smaller company would do better than the larger company in 1 month, the two companies were predicted to do equally well in 6 months, and the larger company was expected to do better than the smaller company in 10 years. In other words, participants believed that the winning streak of the smaller company (vs. the larger company) was more likely to continue only in the short term (for 1 month). Consistent with this pattern of judgments, participants also thought the larger company would have a more stable stock price than the smaller company.

Burns (2003) concluded that people expect streaks produced by a more variable process to continue longer. But, because people think that streaks do not last forever, they hold more predictive power in the immediate, short-term future. He argued that perceived variability in success probabilities leads to increased reliance on streak information because people believe the streaks could indicate a change in the underlying generating process (the base rate). This interpretation is consistent with a model presented by Rabin (2002, discussed later in this article), which also predicts that people are more likely to expect the hot hand for sequences they believe are more variable.

Summary and Comments

The most widely studied judgments of nonrandom sequences are those involving sports performance. While actual sports performance statistics indicate that streaks and serial dependencies occur in only a few sports, people associate (too) many sports sequences with positive recency and see the hot hand where it is not present. People expect nonrandomly produced sequences (such as basketball shooting, salesperson's sales rates, and weather patterns) to exhibit longer streaks and fewer alternations, compared to true Bernoulli sequences. Naive investors' seem to expect stock prices to be streakier than they really are, although the analysis of actual price patterns suggests that momentum or contrarian patterns are very subtle and depend on industry and temporal parameters in a complex manner.

In the remainder of this review, we discuss current theoretical accounts of sequence judgments and then we propose a typology of sequence generation mechanism beliefs. The focus of these accounts has been to explain the dramatic and reliable differences in judgments of very similar sequences occurring in the random mechanism versus skilled performance domains. There is a progression of accounts, from the simplest interpretations in terms of people's capacity to pick up statistical regularities, to prior beliefs about the nature of the processes that generate the events. We conclude that the more elaborate cognitive interpretations are necessary to account for the discrepancies between judgments and actual sequences as well as the differences between the two extensively studied domains.

Theories of Sequence Judgments

It is important to remember that it is not one measurement alone, but its relationship to the rest of the sequence that is of interest. (Deming, 1984, p. 3)

There are several theoretical interpretations that address the question of why people have different expectations for sequences in different domains and why those expectations sometimes systematically depart from reality. We begin with accounts that posit minimal conceptual machinery in their explanations and work up to more elaborate accounts. In general, when people learn to make predictions, there is a tendency to move from judgments based on statistical regularities to simple rules and eventually to rules that are embedded in a conceptual system (e.g., explanations, mental models, folk theories). This trend is seen in many studies of learning concepts, grammars, categories, and, in the present context, judgments about sequences (Harvey, Bolger, & McClelland, 1994; Murphy & Medin, 1985; Schank, Collins, & Hunter, 1986). Thus, it is no surprise that theoretical interpretations can be ordered from those that posit simple and intuitive processes to those that develop fairly elaborate mental representations and complex inferences. Because of the emphasis on gambler's fallacy versus hot hand biases in past behavioral research, we pay special attention to how each theoretical position accounts for the differences between negative recency and positive recency prediction biases.

Rational, Adaptive Models: The Best Objective Model of the Environment (Sequence) Is the Best Model of What Is in the Observer's Head

Some theorists argue that even in the complex situations studied by gambler's fallacy and hot hand researchers, we do not need to rise above the level of statistical regularities in the environment to account for sequence judgments. In essence, these theorists propose that observers adapt effectively to the statistical structures in their environments (Estes, 1964; Wilke & Barrett, in press). Since the mind accurately reflects the statistical structure of the environment, a valid model of the environment is *ipso facto* a valid model of the contents of the observer's mind. For example, Pinker (1997, p. 346) commented that perhaps when a theorist attempts to explain an observer's judgment that a dry spell will end in rain by citing an erroneous gambler's fallacy belief, *the theorist* may be engaged in confabulation. In fact, dry spells must end with rain and the observer is making a valid induction from past experienced sequences of rainy and dry weather (Pinker also provided the more apt statistical example that the 100th car on a train is more likely to be followed by a caboose than the third car on the train; cf. Yackulic & Kelly, 1984). Altman and Burns (2005) and Griffiths and Tenenbaum (2004) also proposed that valid models of information in actual sequences are good candidates for models of an observer's cognitive representations of those sequences.

Given the simplicity of the structure of most binary event sequences, simple stochastic models are sufficient to describe most environmental structures. First and simplest, there are elementary probabilistic generators such as those described as Bernoulli "independent and identically distributed" processes, with fixed base rate probabilities for each event category and no influence of the event on trial n on the probability of occurrence of events on trial $n + 1$. Second, there are probabilistic generators that posit constant dependencies between adjacent events such that $p(A|A) \neq p(A|B)$. Third, there are Markov processes in which simple probability generators are linked to one another so that shifts in base rate

probabilities can occur contingent on immediately prior events. In such a process, the occurrence of Event A might lead to a shift to a new state in which the probability of Event A on trial $n + 1$ is higher than prior to the occurrence of Event A on trial n . These simple models are sufficient to describe most sequential dependencies that occur in actual sequences.

However, there are some types of binary event sequences that seem to demand nonprobabilistic models. For example, many sequences of human behaviors may be generated according to a deliberate cognitive plan by an intelligent agent. When researchers get to such cases, they will need models that represent complex deterministic plans. Here the best candidates would seem to be discrete automata theory models from computer science and linguistics. These models can represent any precisely specifiable plan for sequence generation—such as a plan that balances outcomes (perhaps to serve a goal to produce symmetry or equity in outcomes), or a plan that produces a mirror image sequence, or a strategic sequence intended to deceive an opponent about an agent's next move in a competitive game.

Note that we have listed the most useful models for describing sequences and predicting next events, but we have not addressed the issue of what model, for an agent, will maximize desired outcomes (and minimize bad outcomes). The analysis of rational models is difficult because we need to specify the agent's goals precisely to derive a normative model. What is the agent attempting to maximize (minimize)? In practice such models are difficult to falsify because the agent's incentives and goals are usually difficult to determine, and if the rational model is fitted with the agent's goals as free parameters the model can be treacherously flexible. In natural settings the relevant incentives are always complex and in laboratory settings researchers often fail to provide clear instructions or payoffs such that the relevant consequences are not clearly defined for the participants.

However, there are some situations in which human behavior is anomalous enough that it is difficult to credit agents with even approximately adaptive strategies. For example, observers mispredict (and probably misrepresent) sequence structures in casino games (gambler's fallacy), in their purchase of lottery tickets, and in basketball hot hand judgment errors. Here the defense of the rational approach notes that these are odd environments (e.g., casinos) in which habits that are adaptive in other environments may be misapplied, or they involve complex stimulus events (basketball games) where different aspects of sequences may be confused perceptually or conceptually with one another (e.g., confusing a player's confident style with his actual scoring pattern).

Rabin (2002) made the important point that many judgments demonstrating a negative recency bias are not true gambler's fallacy errors because they do not lead to systematic losses. For example, given that red and black are equally probable outcomes in roulette, it is not maladaptive to bet against streaks—either bet is equally dumb. Also patterned judgments and choices may be part of an exploration (e.g., hypothesis-testing) strategy in a novel environment and therefore be adaptive in a general sense. But, as noted before, an analyst must know a great deal about the judgment situation (environment) and about the agent's goals in order to make a convincing argument for (or against) adaptiveness. For instance, Rabin also cited several examples from lottery and race-

track betting where betting on negative recency expectations is costly and, thus, represents a true gambler's fallacy.

Urn Model Interpretation of the Gambler's Fallacy

An interesting explanation of biases toward negative recency proposes that people reason about apparently random sequences as though they are true random samples from an Ehrenfest urn, but with incorrect assumptions about sampling without replacement. An Ehrenfest urn is an idealized mechanism that produces events (traditionally signified by colored balls) in an independent, haphazard order. Obviously, if one is sampling randomly from an urn that contains only a few balls, and one samples without replacement, a gambler's fallacy pattern would be expected. For example, suppose one samples without replacement from an urn containing two red and two black balls. A draw of one red, without replacement, drops the likelihood of a repetition of red on the next draw from two fourths to one third, a true negative recency pattern. Real-world examples include adaptively significant situations that involve foraging for a reward under conditions where obtaining the reward on one trial drops the probability of obtaining it on the next (e.g., hunting, investing). The sampling without replacement model has been developed most rigorously by Rabin (2002; for precursors, see Fiorina, 1971; R. S. Morrison & Ordeshook, 1975; Rapoport & Budescu, 1997).

Rabin (2002) introduced the urn model and showed that sampling without replacement from small urns easily provides an account for gambler's fallacy type judgments. He acknowledged that fitting the results of specific behavioral studies may not be so easy (p. 787), but in principle the model provides a plausible account for gambler's fallacy judgments. His primary examples are predictions of hypothetical stock prices or analyst forecast successes, and he also reviewed many of the studies from the behavioral literature showing gambler's fallacy and hot hand predictive habits.

Rabin (2002) developed a complementary analysis of overinference from small samples (see also Mullainathan, 2002). The first part of his analysis—sampling without replacement from a small urn—predicts that people will expect reversals. The second half describes a process in which the observer jumps to conclusions about the urn and expects too much streakiness in a sequence. Thus, Rabin's complex model can, in principle, account for both judgments that predict too many reversals (gambler's fallacy pattern) and too few reversals (hot hand pattern).

Rabin's (2002) approach precisely specifies complementary processes that are sufficient to produce one or the other bias, but his analysis is incomplete. First, although the account can generate quantitative predictions, the two-component model has not been fitted to behavioral data sets. Second, additional theoretical principles are needed if the two-component framework is to make a priori predictions of the incidence of gambler's fallacy versus hot hand patterns of judgment. Rabin modestly pointed out that further assumptions will be needed to provide definite predictions of when each bias will appear in actual behavioral data.

Interpretations Based on Folk Theories About Luck and Randomness

Many naive explanations for events in sequences refer to everyday notions about luck and skill associated with games of

chance and sports (Nickerson, 2002; Sundali & Croson, 2006; Wagenaar, 1988). A plausible interpretation is that these beliefs are based on experiences with sequences that contain causally dependent events such as repeated sampling of resources that deplete (and therefore exhibit negative recency) or events that exhibit true causal momentum as in some skilled performances and mechanical devices (producing actual positive recency). People then overgeneralize from sequences in which there are causal dependencies to new situations in which the events are, in fact, independent (cf. Ayton & Fischer, 2004, p. 1376; Yackulic & Kelly, 1984).

There are miscellaneous rules about luck and chance that seem to be different from beliefs about random generators but are probably not consistently shared among adults in Western cultures. For example, Sundali and Croson (2006) found several variant notions including streak-of-luck beliefs leading to positive recency expectations in roulette betting; as well as stock-of-luck beliefs leading to the opposite expectations. Evidence for streak-of-luck beliefs was also found in Ayton and Fischer's (2004) study on judgments of a player's predictions of roulette outcomes. Friedland (1998) determined whether participants were luck-oriented or chance-oriented individuals, and then asked participants to play games in which selecting from different probability decks resulted in different win/loss outcomes. Friedland found that whereas chance-oriented participants did not show recency biases in either of two experiments, luck-oriented participants tended to exhibit belief in the hot hand (by betting more in Game 2 after winning in Game 1) and the "stock-of-luck" (betting less in Game 2 after winning in Game 1). These findings suggest that people who believe in causal luck may be more likely to exhibit recency biases than people who believe in chance (n.b., betting habits, compared at the game level, are a very indirect measure of recency beliefs, and conclusions based on such indicators can be only tentative in this regard).

When experienced gamblers were asked to describe the determinants of outcomes in blackjack and roulette after the fact, they attributed outcomes to three factors: chance (18%), skill (37%), and luck (45%; Keren & Wagenaar, 1985). Wagenaar and Keren (1988) found that college students referred to *luck* to explain difficult desired (intended) accomplishments or escapes from negative consequences, whereas *chance* explained unanticipated personal experiences, viewed as outside the individual's control. Teigen (2005) also noted that many references to luck involve situations in which a bad outcome is experienced but a much worse counterfactual outcome comes easily to mind (as in the apocryphal story of Lucky, the tailless, three-legged, blind dog who repeatedly escaped death). Thus, the concept of luck seems to be invoked as an after-the-fact explanation when unexpected good or bad events happen that the actor would intentionally seek or avoid. Consider, for example, how an astounding sequence of wins or losses at roulette, a novice player successfully making a half-court shot during a basketball game, or an experienced card player who is dealt bad hands all night are likely to invoke attributions to luck.

Wagenaar (1991) proposed three properties for formal random outcome generation devices: a fixed set of alternative outcomes, the selection of outcomes independent of prior outcomes, and equiprobable outcomes. Blinder and Oppenheimer (2008) found that only independence was directly related to college students' use of the term *random*. Tversky and Kahneman (1971) proposed

that people judge random events according to a *law of small numbers*, such that even a very small sample of events (short subsequences) is expected to represent the properties of the population of events (or the conceptual properties of the sequence generator). Bar-Hillel and Wagenaar (1991) provided a summary of naive, informal expectations for random sequences consistent with the law of small numbers: close correspondence between relative frequencies and ideal base rate probabilities in short subsequences, and irregular patterns that appear unpredictable and difficult to conceptualize (notably, sequential independence was not mentioned). When a person confronts a mechanism that she thinks is random, she expects all outcomes to be represented in proportions approximating their expected base rates (even in short sequences) and for the pattern of outcomes to look irregular. Kareev's (1992) insight that the capacity of working memory may provide a clue as to the kinds of short sequences that are expected to be representative of the ideal generated sequence. The law of small numbers provides a good account of many of the results concerning judgments of events produced by coins, roulette, and the sex of births in small samples. This account does not apply to generators that are believed to be nonrandom (e.g., the products of goal-directed behavior by a skilled athlete).

Gilovich et al. (1985) elaborated on the law of small numbers, invoking the representativeness heuristic interpretation to explain the hot hand beliefs of basketball fans. Their account is as follows: (a) Observers begin their observations of basketball play by considering the possibility that players' sequences of hits and misses are generated by a random process; (b) observers' misconceptions of randomness lead them to expect sequences that (over-) represent the base rate probability of hits, that show too many reversals (given the prescription of inter-event independence for a binomial process), and that look unstructured and haphazard; (c) when they observe basketball shooting (which is, in fact, descriptively Bernoulli random), their (erroneous) expectations (about reversals and short runs) are disconfirmed and they infer that the shots are interdependent, streaky, and demonstrate hot or momentum shooting.

The Gilovich et al. (1985) interpretation has been criticized for incompleteness (Ayton & Fischer, 2004; Boynton, 2003; Gigerenzer, 1996). The incompleteness is apparent if one compares coin tosses (where reversals are expected by observers) and basketball shots (where streaks are expected). If both series are binomial, why would not observers be surprised in both cases by the unexpectedly long runs? Why would not they conclude that neither sequence is random (given their technically incorrect law of small numbers beliefs about what a random sequence would look like)? This would imply that observers would react similarly to basketball players ("This guy is not random; he's streaky and hot") and to coin tosses ("This can't be a fair coin; it's too streaky"). But, observers do not react similarly to coins and athletes, so something must be missing from the original interpretation. As Boynton (2003) put it, "If we insist that local representativeness explains both effects, we need to supplement this heuristic with something else that would guide the use of the randomness prototype" (p. 126; see also Ayton & Fischer, 2004, p. 1376).

There is a further shortcoming of the account based on the law of small numbers. In a questionnaire to sports fans, Gilovich et al. (1985) discovered that the fans had intuitive beliefs about proba-

bilistic dependencies among shots (that the probability of a hit following two immediately prior hits was greater than following two misses). Roney and Trick (2003); Benhsain, Taillefer, and Ladouceur (2004); and Ladouceur et al. (1996) have also found evidence that observers believe in causal dependencies between events in sports and in casino game outcomes. But, the simple form of the law of small numbers interpretation merely says that observers will expect small subsequences to exhibit the properties of the prototype sequence. Causal dependency is not part of the prototype. A careful reading of the original Gilovich et al. article suggests that the authors realized that additional conditions were necessary to distinguish between coins and basketball. They noted that

A major difference between the two processes [coins and basketball] is that it is hard to think of a credible mechanism that would create a correlation between successive coin tosses, but there are many factors (e.g., confidence, fatigue) that could produce positive dependence in basketball. (p. 313)

They also noted that there are other factors that could enhance the impression of dependencies (hotness or coldness) in basketball such as the style of play and other features of behavior besides merely the pattern of successful shots. A further dissection of fans' beliefs about hotness, coldness, and related concepts seems warranted.

To its credit, the law of small numbers account does address the difference between the gambler's fallacy and the hot hand, and it provides the beginning for a cognitive interpretation. Several researchers have also emphasized perceived randomness of the outcome generators in expectations about sequences: Burns and Corpus (2004); Gronchi and Sloman (2008), and Tyszka et al. (2008) have said that random versus nonrandom is the key distinction; Ayton and Fischer (2004) said it is random versus skilled, or inanimate versus human. But others such as Caruso and Epley (2008) have suggested that the primary distinction is between intentional versus unintentional generators (see also Choi et al., 2003). There is a focal disagreement here on the nature of the conceptual dimension that drives differences in expectations. Tyszka et al. (2008) provided direct results on this issue, finding that judgments for coin tosses and fortune-teller predictions are similar, while judgments for weather and basketball are similar—casting some doubt on whether intentional/unintentional is the key difference. (We revisit these distinctions in more detail later.)

Perhaps more important is that all of these researchers addressed the gap in the law of small numbers account. They all appeared to reject the notion that observers begin with the default assumption that all sequences are random and then shifted to an alternative interpretation when their naive conception of randomness was violated. Rather, observers began with an expectation of what the sequence should be representative of (e.g., expecting nonrandom, controlled streaks if they are watching a basketball player and expecting [too] many reversals if they are watching a roulette wheel). According to Ayton et al. (1989), "before any observation can occur then a particular theoretical orientation must be adopted" (p. 223).

Complex, Cognitive Mental Models of Sequence Generators

It is productive to take the notion of heuristic explanations one step further and propose that in most situations in which sequence judgments are made, people have a collection of interpretative beliefs or a mental model ready in mind that allows them to comprehend information about sequences, infer properties of sequences, and make predictions of what will happen next. We call this approach *explanation-based* because the systems of beliefs people use to make such inferences are usually smaller and less orderly than theories, but larger and more integrated than a simple heuristic or rule of thumb (Hastie & Pennington, 2000). When people encounter a new, interesting, important series of events, one of the primary comprehension goals they have is to explain how it works and they use systematic beliefs, stored in long-term memory, to create sense-making explanations.

A simple scenario for explanation-based predictions of events in a sequence would take the following form. A person is observing a series of events that is perceived as a sequence because they seem to be generated by a single mechanism—for example, watching one basketball player shooting baskets in a college game or watching one roulette wheel spin in a casino. The person has prior experience with events of that type or something similar (a basketball fan, a recreational gambler); therefore, he or she has prior beliefs about determinants of the events and these beliefs will be retrieved from the observer's long-term memory to make sense of what's happening (to explain why a shot was missed, why the ball bounces into a red slot, etc.). When the observer wants to predict the next event in a sequence (the chances that a player will score, the chances a color will be selected), he or she considers the recent events in the sequences (the last four attempts were successful [were red]) and relies on his or her background beliefs to infer what will happen (a belief that a player is skillful enough to exert control over the ball, that black is now due). These background beliefs can be loosely described as a mental model of the mechanism that is producing the outcomes.

It is important to note that most predictions of "What's next?" in a sequence will be a mixture of top-down reasoning (the observer enters the situation with definite beliefs about the generating mechanism) and bottom-up context sequence information. These inferences can be complex; for example, consider an observer watching a roulette wheel. He or she is likely to bring to mind a rough mental model of how such devices produce their outcomes (naive mechanical principles and erroneous beliefs about random processes). Then, when a run of reds is observed (e.g., four in a row), the context sequence plus the mental model of the random mechanism lead to the inference that the wheel will reverse and come up black on the next spin. If the context sequence had been a haphazard series of reds and blacks, or a run of 20 blacks in a row, the general tendency to predict a reversal would be attenuated. (Note that a typically socialized adult, who knows something about casinos, is unlikely to walk into the casino and begin testing the hypothesis that the wheel is a random device. Although, if the sequence generated by the wheel is unexpected [e.g., a run of 20 blacks, a run of 20 perfect red/black alternations], then the observer might reconsider his or her initial context-dependent belief that the wheel is a random generator.)

Let us consider the basketball case as well. Here a typical fan walks into the arena with definite ideas about how basketball players work—players are causal, intentional agents with substantial control over outcomes and a simple goal of hitting baskets. Most fans also have additional beliefs about motivation in skilled performance, including the notion of hotness and coldness. (These beliefs may be very specific, e.g., that hotness and coldness apply to only some sports, that some players or teams show more streaky behavior than others, etc.) When a player exhibits a distinctive pattern in performance (e.g., a run of four successes in a row), the fan will rely on his or her beliefs about the player (mechanism) plus the information in the context sequence to infer something about the situation or to predict the next event. In the case of basketball, a sequence of successes is likely to lead to predictions of more successes, perhaps based on the notion that intentional, skilled, goal-directed events involve momentum or other causal dependencies. (Again, note that fans do not enter the arena with the default hypothesis that a player's performance will be random.)

Summary and Comments

Past theoretical accounts of sequence perception include those based on statistical regularities in the sequence, an urn model sampling-without-replacement interpretation, and heuristic folk theories (such as the law of small numbers account of randomness). These models are somewhat limited in their ability to explain why or when sequence perceptions diverge from reality, or when predictions for different sequences will tend to exhibit positive versus negative recency.

Our review of prior research suggests that judgments about the sequence are based on beliefs about the generator, knowledge of recent events in the sequence, and other contextual factors (e.g., other conditions, states, or influences that are present in the situation in which the sequence is encountered). We assert that most of the research involves situations in which people are deliberately (consciously) thinking about the prediction task and are relying on previously learned explanations or mental models of the generating mechanism. People have a rich store of beliefs and rough folk theories that they rely on heavily in making sequence predictions. Thus, our inclination is to develop a mental models, explanation-based interpretation of the cognitive processes underlying predictions of "What's next?"

An Explanation-Based Account of Sequence Judgments

To flesh out an explanation-based account of judgments about events in binary sequences, we need to provide some sense of what those mental models look like (for similar exhortations, see Alter & Oppenheimer, 2006; Moldoveanu & Langer, 2002). One of the major obstacles to successful theoretical analyses of sequence judgments is the lack of a systematic representational scheme in which to express the mental models. It is no accident that mental models have been most studied in domains involving geometric and geographical relationships and simple mechanical devices where familiar, useful representational systems are readily at hand. In the remainder of this review, we propose a systematic theoretical framework in which to represent and analyze them.

Observers' Beliefs About Sequence Generators

The two most-studied cases of sequence predictions, random devices and skilled athletic performance, are associated with specific expectations and specific mental models of the event-generating process. For these two cases, we can tender hypotheses about the prediction habits of observers based primarily on empirical generalizations from the past studies reviewed above. We can also venture beyond these two well-studied cases and try to conceptualize a framework that is useful for modeling other types of sequences. Several perceived properties of the generating mechanisms (randomness, intentionality) have been repeatedly mentioned in theoretical discussions of gambler's fallacy and hot hand beliefs. Other properties (control, goal complexity) are based on our interpretation of the research synthesis and our own experimenter intuitions. We believe that the extent to which randomness, intentionality, and control are perceived by the observer as descriptive of the sequence-generating mechanism is strongly indicative of the observer's predictions of positive versus negative recency. We also suggest that beliefs about the complexity of the sequence mechanism's goals can moderate those prediction tendencies. Note that in our view, they are continuous, rather than dichotomous, dimensions.

Randomness

It is apparent that people have a concept of a random mechanism and this concept is widely shared across educated adults in industrialized societies. Most college students believe that birth sequences, casino devices (dice, cards, roulette wheels), and many mechanical processes (coin tosses) are random. When they encounter a novel device and are unable to predict the events it generates, they are likely to deem it to be random. People generally expect random mechanisms to produce sequences with short runs, negative recency, and a representative base rate. However, some random processes can lead to positive recency beliefs (e.g., believing that roulette players, or numbers, can get hot) and some nonrandom processes can elicit negative recency thinking (e.g., believing that it cannot rain forever and the streak of rainy days will end). Burns and Corpus (2004) found that participants' randomness ratings for different scenarios correlated only 0.29 with statistical properties of their predictions. Thus, randomness does not appear to be the only factor driving people's expectations.

Intentionality

As we discussed earlier, several theorists have argued that intentionality is a major consideration in expectations for streaky performance by human agents (Ayton & Fischer, 2004; Burns & Corpus, 2004; Caruso & Epley, 2008; Choi et al., 2003). The more intentional and goal-directed a sequence-generating mechanism is perceived to be, the greater the tendency to expect positive recency. For example, Caruso and Epley (2008) found that people were likelier to expect streaks to continue when an agent producing the same sequence of outcomes was intentional (a human) rather than unintentional (a robot) and when intentionality was highlighted by instructions to observers. In a study by Ayton and Fischer (2004), intentional roulette players exhibited hot hand thinking when con-

sidering the success of their roulette predictions but simultaneously exhibited the gambler's fallacy when asked to predict the color of outcomes generated by the unintentional roulette wheel.

Control

Some events are more controllable than others by the intentional agent producing the outcomes. The grades a student gets in school are more controllable than the profits from an investment, which are in turn more controllable than the winnings from playing lotteries; for skilled players, golf putting is more controllable than basketball field goal shooting, which is more controllable than baseball batting. Professional athletes exert more control over outcomes than amateur or recreational players. Sequences involving more highly controlled outcomes elicit a tendency to expect positive recency.

The extent to which sequential outcomes are controllable correlates positively with the length of streaks that actually occur in observed sequences. For example, Gilden and Wilson (1995) showed in a lab environment that streaky, hot hand performance in golf and darts depends on the skill level of the player (i.e., how much they could control outcomes). Unusual performance streaks were found only for players who had moderate and high hit rates and not for players with lower hit rates. The degree of control that athletes have over outcomes in certain sports may explain why researchers and statisticians have found evidence for actual hot hand performances occurring in individual sports such as bowling, archery, horseshoes, and billiards but not in more chaotic team sports such as basketball, baseball, and volleyball.

When it comes to judgments of sequences of events relevant to purposeful behavior, researchers expect the perceived intention of the agent to interact with the perceived controllability of the outcomes (see Malle & Knobe, 1997, for a general discussion of people's folk concept of intentionality). Research on achievement attribution theory is consistent with this idea, finding that people attribute achievement success and failure to four causal factors—luck, effort, ability, and task difficulty—and the latter three are associated with the level of one's control over the outcome (Frieze & Weiner, 1971). To illustrate how outcomes can depend both on intentionality and controllability, consider (again) human performances in sports. An athlete's intentions to score points are constant during a competition, but there are likely to be fluctuations in the athlete's ability to carry out his or her intentions due to the actions of opponents and teammates, changes in playing conditions, fatigue, and so forth. (Of course, ability or skill depends on several things, such as the person's innate talent and training.) No matter what their intentions, none of the authors of this article is going to dunk the ball in a professional basketball game, and no observer would expect them to.

Goal Complexity

We hypothesize that the perceived complexity of an agent's goals will moderate expectations of positive and negative recency. When simple goals are involved in binary event sequences, one outcome is unequivocally more desirable than the other—for example, in most sports, players have one simple goal: score. For the simplest of goals, this is true regardless of

a player's personal degree of control over the outcomes or whether the playing conditions are changing over time—one outcome is always good (e.g., hits); the other is always bad (e.g., misses). When an agent with control over the outcomes pursues simple goals, we believe observers will tend to expect positive recency and streaky sequences.

But goals can also be strategic and contingent; this usually occurs when two agents interact with one another to mutually determine the relevant outcomes. A tennis player may have a simple overriding goal, to score points. But, in deciding whether to place the ball on the opponent's left or right side of the court, the player invokes a more strategic goal. In this case, the subgoal (of ball placement) to achieve the ultimate goal (win the point) is variable and depends on the opponent's goals. At any given point in the sequence, whichever of the two possible outcomes (left or right) is the desired outcome can depend on what the previous outcomes were, on external influences such as the opponent's reactions and skills, and on the player's overall repertoire of strategies for outcome patterns (e.g., "I want to make the opponent run back and forth to tire her out," "Her backhand is weak, so . . .," or "I will try to play as unpredictably as possible and make it impossible for her to anticipate me"; cf. Walker & Wooders, 2001). To the best of our knowledge, no research has been done on how goal complexity affects sequence judgments. We believe that the simple bias to expect more frequent and longer streaks for intentional, controlled processes is attenuated when the relationship between goals and outcomes is complex.

We now have a good understanding of the mental models for the two most-studied cases of sequence judgments: a random mechanism model and an intentional, skilled (controlled) mechanism with a simple goal. This leaves several other common but poorly understood cases, most notably: nonrandom but unintentional mechanisms such as those believed to underlie sequences generated by many mechanical devices (e.g., whether your personal computer crashes when you open a PDF file), natural phenomena (rain, hurricanes, temperature changes), biological organisms (growth of a plant), and fluctuations in the stock market. For these generating mechanisms, we speculate that observers will expect occasional short streaks, trends, or cycles (indicating nonrandomness) depending on the specific generating mechanism and on information they have about recent outcomes.

Another important class of sequences is produced by intentional agents with relatively less control over sequence outcomes (e.g., your kid sister's first attempt to play basketball, the outcomes experienced by an ordinary poker player, and success in predicting the stock market by a typical analyst). Here we conjecture that skill or control is the key moderator of the length of streaks expected to occur—more control, longer streaks.

Another proposed class of sequences consists of those produced by agents with control over outcomes and strategic complex goals. Sequences of this type include decisions or choices for which one outcome is not always better than the other (e.g., left/right placement of tennis serves or soccer penalty kicks). In cases when the outcomes are controllable, sequences are determined by the intentions of the sequence generator. In other words, if a person has high control over the outcomes, the outcomes will be whatever the person wants them to be. (Consider a hypothetical scenario in which a son must decide which divorced parent to visit each Christmas—Mom or

Dad. He could consider the parents' feelings and alternate visits, or base the decision on current weather conditions, or on the fact that he hates Dad's new girlfriend.) Therefore, it is difficult to predict the general properties of these types of sequences, though we hypothesize that judgments for sequences involving a strategic goal are less likely to exhibit positive recency than sequences involving a simple goal.

Because the situations in which sequences are generated are often complex and multicausal, we believe in many cases people will attribute sequence outcomes to a mixture of the four model characteristics we just described. Organizing the taxonomy of generator models as Ballantine ring diagrams would emphasize all of the possible combinations of characteristics. However, in Figure 1 we present an example representation of the types of mental models for sequence generators, based on the proposed four characteristics, as a classification tree graph. This simplification allows us to more clearly organize sequence generators as one of the hypothesized five types of explanatory models we believe are the most commonly encountered in the real world. In our example representation, the four binary dimensions can be summarized as a lop-sided classification tree, with 16 potential cells, but with only five occupied by plausible mechanisms: i. Random, ii. Nonrandom Unintentional, iii. Intentional With Less Control, iv. More Control With Simple Goal, v. More Control With Strategic Goal.

We should note that, despite the hierarchical structure of Figure 1, we do not mean to suggest that consideration of the four dimensions occurs in any particular sequential order. We put randomness at the top in Figure 1 because of its prominence in past research. It is possible that, unless the situation is truly novel and mysterious, people most often enter a situation with a definite hypothesis about the mechanism (and only when people cannot determine anything about the causal mechanism do they place a sequence in the random class).

Summary and Comments

We propose that there are four perceived properties of the sequence-generating mechanism that are important for characterizing people's mental models of binary sequences: randomness, intentionality, control, and goal complexity. We argue that while random mechanisms lead to a tendency toward gambler's fallacy/negative recency predictions, more intention, more control, or simpler goals are each associated with a tendency toward hot hand/positive recency predictions. An example conceptual framework is presented for organizing and classifying mental models of common sequences according to these dimensions.

Figure 1 serves as an illustration of how an explanation-based typology—one that goes beyond heuristic, unidimensional accounts of sequence judgments—might be conceptualized. Our classification tree does not attempt to cover the entire space of all possible relevant dimensions, sequence generators, or sequence-generating situations. Rare, artificial, or particularly complex generators may not be easily classified. Nor do we mean to suggest that every person would classify every sequence generator similarly (although some sequence generators may result in less variable classification than others). Variability in the selection of explanatory models may stem from

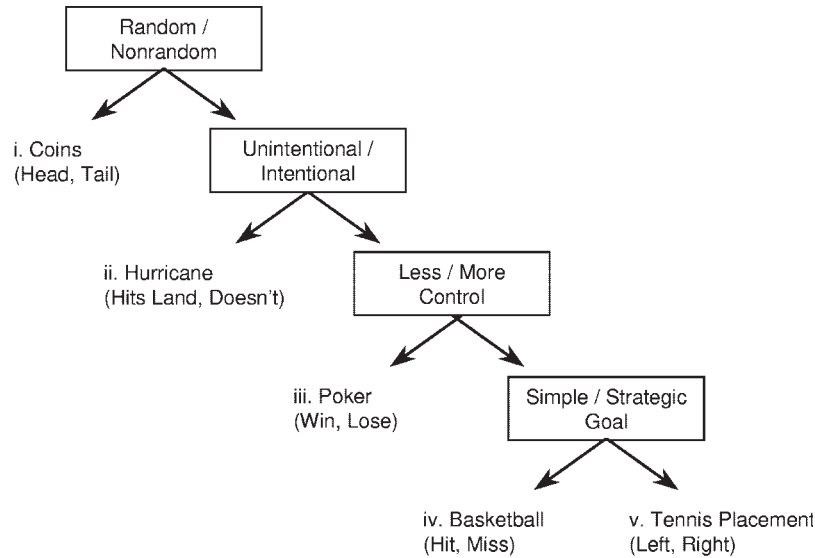


Figure 1. An example classification tree for mental models of sequence generators, based on perceived characteristics of the generator: the randomness and the intentionality of the mechanism, controllability of outcomes, and complexity of the mechanism's goals. (Note that in our view, they are continuous, rather than dichotomous, dimensions.) Example scenarios for five classes of sequences are provided.

differences in people's previous experience, intuitions, various personality variables, and so on. Gamblers may believe that gambling outcomes are more controllable than nongamblers. Meteorologists may believe that the weather is less random than laypeople, and certain tribal populations may believe in intentional weather gods controlling the weather. Although there may be individual differences in which mental model people select for each sequence generator they encounter, we contend that the proposed basic set of mental models may very well adequately represent the majority of sequence generators relevant to everyday experience. In the next section, we propose the Markov model approach for describing the properties of a specific sequence as well as people's specific perceptions of that sequence.

Models for Representing Actual and Perceived Binary Sequences

Only three types of systematic patterns are cited repeatedly in efforts to model empirically observed sequences of binary events: (a) trends up or down in the base rates of occurrence of the two outcome types, (b) regular cycles of gradual changes up and then down in the rates of occurrence, and (c) multistate shifts in base rates (e.g., hot vs. normal performance). Of course, each pattern is described (and interpreted) in terms of detailed domain-specific causal mechanisms. For example, a trend in the outputs of a mechanical system might be described in terms of forces, momentum, energy dissipation, or wear and tear; in a biological system in terms of growth, development, disease, erosion, or depletion; and for single-agent behavioral events with reference to learning, motivation, fatigue, forgetting, or laziness.

The traditional approach is to model binary event sequences in terms of Markov process stochastic models. The simplest case is an unpatterned, Bernoulli independent and identically distributed random sequence, like the sequences observed in coin tosses

(head/tail), human births (male/female), and basketball scoring (hit/miss). The stochastic mechanism that generates such a sequence is usually represented as a (nonhidden) Markov process, as in the left panel of Figure 2. The circles represent observable outcomes (e.g., h/t for head/tail), and the arrows represent the probabilities of transiting from one state to another—all .50 in the case of fair coins and idealized human births (the actual transition probabilities are not exactly .50 for male/female births). Most of the actual sequences studied in research on judgments of "What's next?" are best described by this Bernoulli process generator, and the interesting finding is that observers see patterns and inter-event dependencies where there are none.

Hidden Markov models are the standard approach to modeling time series data for discrete events (e.g., binary event sequences) in linguistics, engineering, physics, and biology. Essentially, they provide a probabilistic mechanism that generates events in a sequence. The general form of such models is presented in Figure 3 where the relationship between hidden underlying causal generators (usually represented as probabilistic Ehrenfest urns) and the sequence of observable events is depicted graphically. Incidentally, these are the Markov process models that have been used most extensively in psychology to model human cognitive learning processes (cf. Atkinson, Bower, & Crothers, 1965; Krantz, Atkinson, Luce, & Suppes, 1974). In those early applications in psychology such as a cognitive concept attainment task, the hidden states were usually cognitive capacities (e.g., the participant had learned, partly learned, or guessed a response), the observable events were behavioral achievements (e.g., correct, incorrect responses), and the models generated modal response sequences.

Trends and cycles are easy to model with hidden Markov processes as illustrated with the graphical models in Figure 4. Each could be the model for an actual sequence generator or for the mental model that an observer relies on to make judgments

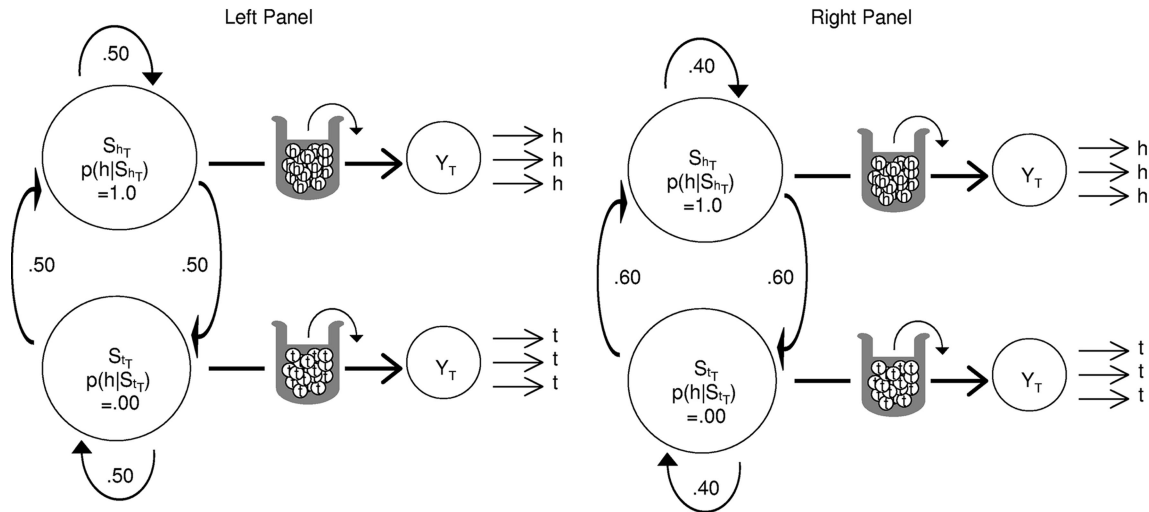


Figure 2. The left panel illustrates a conventional Markov process model for an ideal fair coin; the right panel illustrates a model that mimics the expectations associated with the concept of a random coin. Note that neither model involves hidden processes—the states in the model are observable conditions (whether the observable outcome of a coin toss is a head or a tail). A complete specification of the process would include a vector specifying the initial probabilities of being in each possible state at the start of the process. We assume the probabilities of being in the head (S_h) or tail (S_t) states are equal to .50 in this example. T = time; t = tail; h = head; Y = observed event.

of what event will occur next. Figure 4A shows how momentum can be added into a simple trend. Note that the probability of generating a *hit* (h) increases and decreases across the hidden states in the top half of the diagram and in the bottom half of the diagram. At the same time the probability of transiting into the next state increases, producing a momentum effect (up or down). Figure 4B is a simple model for a regular cyclic pattern in base rate probabilities. In this illustration, transition to the next hidden state occurs in only one direction (but is not certain on any trial, $p(\text{transition}) = .80$), while the

states generate successes (s) with an increasing and then decreasing probability (ranging from .90 at the zenith of the cycle to .10 at its nadir).

In Figure 4C, we describe a hidden Markov model for a three-state system like the one that is often proposed to describe baseball and basketball scoring for a player believed to be hot, normal, or cold. The states are subscripted to represent hot (H), cold (C), and normal (N) (rather than with the time index (t) in Figure 3). Assume the player begins a game in S_N (normal) and is likely to remain in that state from trial to trial across time

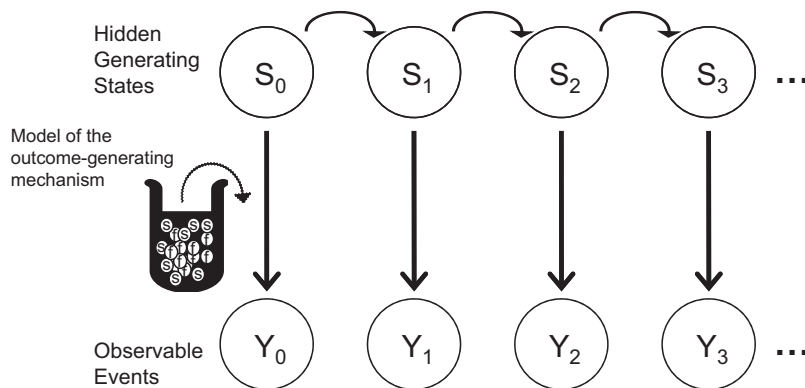


Figure 3. The general schema for a hidden Markov process model: Hidden states (S_t) determine observed events (Y_t), where $t = \{0, 1, 2, 3, \dots\}$, through binary outcome-generating mechanisms represented as probabilistic sampling from urns of outcomes. The outcome-generating urns are usually hypothesized to change composition from hidden state to hidden state, producing shifts in the probabilities of the observable event outcomes (binary and symbolized by s and f in the diagram to indicate generic success and failure outcomes, respectively). Arrows connecting the hidden states represent the probabilities of changing from one state to another or staying in the same state across time (with time passing from left to right).

($p[\text{remain in } S_N] = .80$). But, there is a possibility that the player will shift into a hot (S_H) or cold state (S_C). Once hot or cold, a player is likely to remain in that state ($p[\text{remain}] = .80$, for both S_H and S_C). Each state is associated with a second process that produces observable responses (hits and misses when shots are

made in a basketball game in this example), often represented as sampling from an urn. In this example, we propose that when in the normal state, field goal attempts succeed with a probability of .55 (e.g., sampling with replacement from an urn with 55% hit balls); when in the hot state, succeeding with a probability of .85; and when in the cold state, succeeding with a probability of .25.

In all of these models, the hidden states compose a Markov chain, meaning that the states are discrete and that there is path independence. Path independence (sometimes referred to as the *Markov condition*) means that the current state is only related to the immediately prior state (i.e., given the value of S_{t-1} , the value of S_t is independent of all states prior to $t-1$). This is the type of multi-state process model that has been fitted to baseball hitting data by Albert and his colleagues (Albert & Bennett, 2001; Albert & Williamson, 2001).

Speaking more informally, there are three parts to the overall system: (a) a principle to determine which of several hidden states is the initial active state when the process begins (usually represented as a vector of probabilities that sum to unity; e.g., the basketball player has a probability of 1.00 of beginning the game in a normal state). (b) A set of hidden states that determine the behavior of the system. The system begins in one of these states (e.g., the basketball player starts in normal) and a matrix of probabilities then determines whether the system shifts from one hidden state to another from trial to trial (e.g., the probability that a player who is normal now will become hot on the next trial, the probability that a player who is hot now will become normal on the next trial, etc.). (c) A second set of processes, usually represented stochastically (e.g., sampling from urns) that produces visible responses from each hidden state (e.g., an urn that generates hits and misses representing the player's performance when hot or normal or cold). The Markov model framework illustrated above is the best approach available to describe the actual structure of binary event sequences, and we recommend that anyone conducting studies of the perception of events in a sequence describe the type of process that generates the to-be-judged events in this terminology.

As we mentioned earlier, this Markov models approach can also be used to model the *perceived* structure of the sequence. One route to a systematic description of observers' mental models for sequence generators is to begin with the models of actual sequence generators and then to stipulate the specific aspects of those ideal generators that are nonstandard in the

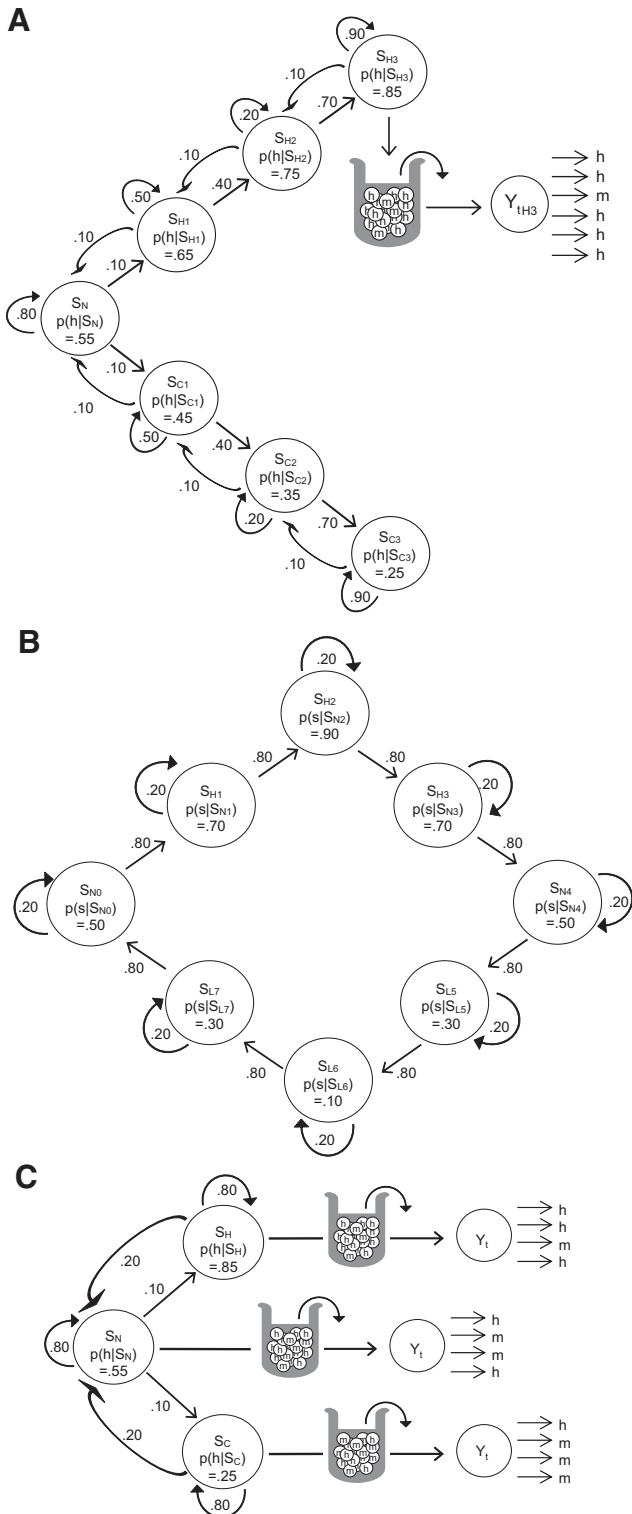


Figure 4 (opposite). Hidden Markov models for representing binary event sequences. Panel A illustrates a trend process (with momentum), and Panel B shows a cyclic process. Note the response-generating urn processes are abbreviated (Panel A; the observable events, h and m , are sampled from the urn associated with each state—only one urn is depicted in the diagram) or omitted (Panel B). Panel C shows a multistate process of the type proposed to represent a basketball fan's judgments of a player's performance, with three hidden states (circles), three response-generating mechanisms (urns), and the observable responses (h , m). Our conjecture is that the fan imagines that the player might be in one of three hidden states that determine the level of his or her performance: normal (N), hot (H), or cold (C). Again, for simplicity the initial state vector is not presented (a reasonable assumption is that the fan would always begin by expecting that the player is in the normal state). Y = observed event; S = state; h = hit; m = miss; t = time; L = low; s = success.

mental models. The right panel in Figure 2 illustrates a model to represent the process in the mind of a person observing a sequence of coin tosses. The .60 transition probabilities from h to t and t to h are consistent with the common tendency to expect too many reversals. This is essentially the tactic followed in studies of mental models of geographical maps and mechanical devices where the model of the actual system (map, device diagram) is also the description of the mental model. How would this tactic be applied to behavioral data? First, a sample of the actual sequences of events would be modeled as hidden Markov processes (ideally the same sequences that will be stimulus sequences for human observers' judgments). Second, human participants would make judgments of a related sample of sequences. Third, the human judgments would be modeled by the same family of hidden Markov process models. Finally, the two models would be compared to determine whether the human behavioral model is different from the actual generating process model. Budescu (1987) provided an exemplary analysis for the simple case of predicting binary events generated by a Bernoulli process like a coin toss.⁴ However, we expect that this ideal application will be rarely achieved. The major problem is that very large samples of judgments would have to be obtained from individual observers to support a systematic quantitative modeling strategy. Nonetheless, we hope that full quantitative modeling of both actual stimulus sequences and observers' judgments will occur in some cases. Realistically, what we strongly recommend is the use of the Markov notation, including diagrammatic conceptual models, to specify researchers' hypotheses about the beliefs about generating mechanisms that they attribute to the observers who render "What's next?" judgments.

What are some differences between the model of the actual sequence and the model that mimics the human's judgment sequence? First, we know that in many contexts (e.g., sporting events), human's will "see" more interevent dependencies than are actually present in the actual event sequence. Second, human judges will rely on recent event sequences to infer which hidden generating state the system is in at any point in time. So, while the ideal Markov process will move from one underlying hidden generating state to the next as a stochastic function of only its current state, humans will infer the new state based on the recent observed events (and their beliefs about the current state). In the example case of the multistate model for a basketball player (see Figure 4C), people rely on the data (a player hits three in a row) to infer which state (hot) the player is in.⁵ Note that this is not exactly analogous to a statistician searching for a best fitting model; rather, it is as though the analyst has selected a model and is attempting to determine the state of the model, within the model framework. This is also a statistical inference problem, and Burns (2004) provided a discussion of how valid inferences can be made from observed streaks to infer hidden underlying states.

Summary

We propose that theorists who want to describe mental models of causal mechanisms that generate binary events in sequences follow our suggestions to specify two Markov process

models: first, a model that will generate the actual observed sequence of stimulus events, and second, a model that mimics the human observers' judgments of what events will occur in the sequence—predicted, remembered, or hypothetical. In behavioral research, it will often be possible to fit models to the stimulus event sequences, as illustrated in the analyses of sports performance data sets reviewed above. However, there will be only a few situations in which enough individual data will be available to formally fit such models to judgments of "What's next?" We still believe that merely adopting the vocabulary and the graphical representations (illustrated in Figures 2, 3, and 4) will provide a substantial increase in precision and discipline of theoretical discussions of the mental models of processes underlying sequence judgments. (The modest sophistication necessary to use this notation qualitatively is readily available in tutorial papers such as Rabiner & Juang, 1986, and Ghahramani, 2002.) Furthermore, stating the conceptual models as graphical Markov process diagrams will allow other, more sophisticated modelers to derive generative implications (hypotheses) and analytically fit judgment data to test the proposals.

Conclusions

Human beings are built to see patterns in sensory and conceptual data of all types (Gawande, 1999; Gilovich, 1993). The capability to induce patterns and to predict what is hidden, what is missing, and what is next is one of our species' greatest achievements and advantage over other animals. The present article provides a review of people's capacities and biases when predicting what will happen next in temporally ordered sequences of binary events. To date, this stream of research has been dominated by studies of two types of sequences: events generated by putatively random devices such as casino games and births and events generated by skilled athletes. Most studies have focused on the differences in predictions for the two types of sequences, a gambler's fallacy bias for random events and a belief in hot hand streaks for sports events.

Despite the narrow focus, studies of judgments of random processes and skilled performances have been very informative. The first lesson is a reminder of the necessity of studying the structure of the to-be-judged sequences. Most observers believe there are streaks in basketball shooting, but statistical analysis has found no evidence for such patterns. Second, simple accounts of the judgment process in terms of learning statistical regularities or the representativeness heuristic seem incomplete as explanations of the between-domains differences in expectations and the departures from accuracy. Our conclusion is that a satisfactory account of these differences implicates the need for a cognitive mental models approach. The mental models account is based on the most fundamental insight of the cognitive approach: People's models of the world, although derived from sensory information and cultur-

⁴ Of course, a Bernoulli process model is not a hidden process model, as it is defined on the observable states only.

⁵ Carlson and Shu (2007) have argued that there is something special about the observation of three similar outcomes in a sequence, that three in a row prompts people to see a streak and infer the current state (hot or cold) from that data.

ally transmitted belief systems, are not always perfect reflections of the outside world. To understand and predict the details of behavior, one needs to understand the cognitive representation of the outside world. We believe that when there are more studies of judgments in domains other than random mechanisms and sports, the need for an explanation-based, mental models account will become even more obvious (cf. Alter & Oppenheimer, 2006).

Finally, we believe that the major obstacle to a successful mental models account is the lack of a systematic terminology to describe sequence generating mechanisms. We propose the hidden Markov models framework as a useful theoretical notation that provides clear and precise descriptions of the mental models underlying judgments. The Markov framework also provides the best statistical models of the actual sequences being judged. We believe this recommendation will lead to further advances in researchers' understanding of the structures of actual binary sequences and the manner in which humans' cognitive systems comprehend and reason about those sequences.

References

- Adams, R. (1992). The "hot hand" revisited: Successful basketball shooting as a function of inter-shot interval. *Perceptual and Motor Skills*, 74, 934.
- Adams, R. (1995). Momentum in the performance of professional tournament pocket billiards players. *International Journal of Sport Psychology*, 26, 580–587.
- Albert, J. (1993). A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association*, 88, 1184–1188.
- Albert, J., & Bennett, J. (2001). *Curve ball: Baseball, statistics, and the role of chance in the game*. New York: Copernicus.
- Albert, J., & Williamson, P. (2001). Using model/data simulations to detect streakiness. *American Statistician*, 55, 41–50.
- Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association*, 88, 1175–1183.
- Alter, A. L., & Oppenheimer, D. M. (2006). From a fixation on sports to an exploration of mechanism: The past, present, and future of hot hand research. *Thinking & Reasoning*, 12, 431–444.
- Altmann, E. M., & Burns, B. D. (2005). Streak biases in decision making: Data and a memory model. *Cognitive Systems Research*, 6, 5–16.
- Anderson, N. H. (1966). Test of a prediction of stimulus sampling theory in probability learning. *Journal of Experimental Psychology*, 71, 499–510.
- Anderson, N. H., & Whalen, R. E. (1960). Likelihood judgments and sequential effects in a two-choice probability learning situation. *Journal of Experimental Psychology*, 60, 111–120.
- Arie, S. (2005, February 11). No 53 puts Italy out of its lottery agony. *The Guardian*. Retrieved September 1, 2005, from <http://www.guardian.co.uk/italy/story/0,12576,1410701,00.html>
- Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). *An introduction to mathematical learning theory*. New York: Wiley.
- Ayton, P., & Fischer, I. (2004). The hot hand and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, 32, 1369–1378.
- Ayton, P., Hunt, A. J., & Wright, G. (1989). Psychological conceptions of randomness. *Journal of Behavioral Decision Making*, 2, 221–238.
- Badarinathi, R., & Kochman, L. (1994). Does the football market believe in the "hot hand"? *Atlantic Economic Journal*, 22, 76.
- Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of "hot hand" research: Review and critique. *Psychology of Sport and Exercise*, 7, 525–553.
- Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12, 428–454.
- Bayer, D., & Diaconis, P. (1992). Trailing the dovetail shuffle to its lair. *Annals of Applied Probability*, 2, 294–313.
- BBC News. (2005, February 11). *Number 53 brings relief to Italy*. Retrieved September 1, 2005, from <http://news.bbc.co.uk/2/hi/europe/4256595>.
- Benhsain, K., Taillefer, A., & Ladouceur, R. (2004). Awareness of independence of events and erroneous perceptions while gambling. *Addictive Behaviors*, 29, 399–404.
- Bennett, D. J. (1999). *Randomness*. Cambridge, MA: Harvard University Press.
- Ben-Porath, Y., & Welch, F. (1976). Do sex preferences really matter? *Quarterly Journal of Economics*, 90, 285–307.
- Blinder, D. S., & Oppenheimer, D. M. (2008). Beliefs about what types of mechanisms produce random sequences. *Journal of Behavioral Decision Making*, 21, 414–427.
- Boynton, D. M. (2003). Superstitious responding and frequency matching in the positive bias and gambler's fallacy effects. *Organizational Behavior and Human Decision Processes*, 91, 119–127.
- Brown, W. O., & Sauer, R. D. (1989). Does the basketball market believe in the "hot hand"? Comment. *The American Economic Review*, 83, 1377–1386.
- Budescu, D. V. (1987). A Markov model for generation of random binary sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 25–39.
- Budescu, D. V., & Rapoport, A. (1994). A subjective randomization in one- and two-person games. *Journal of Behavioral Decision Making*, 7, 261–278.
- 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* (pp. 233–239). Hillsdale, NJ: Erlbaum.
- Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand." *Cognitive Psychology*, 48, 295–311.
- Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand." *Psychonomic Bulletin & Review*, 11, 179–184.
- Camerer, C. F. (1989). Does the basketball market believe in the "hot hand"? *American Economic Review*, 79, 1257–1261.
- Canfield, R. L., & Haith, M. M. (1991). Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology*, 27, 198–208.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52, 57–82.
- Carlson, K., & Shu, S. B. (2007). The rule of three: How the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes*, 104, 113–121.
- Caruso, E., & Epley, N. (2008). *The road to streaks is paved with intentions: Perceived intentionality in the prediction of repeated events*. Unpublished manuscript, University of Chicago, Center for Decision Research.
- Chang, K. (2003, October 22). A 7-game world series is unusually common. *The New York Times*.
- Cheng, L. T., Pi, L. K., & Wort, D. (1999). Are there hot hands among mutual fund houses in Hong Kong? *Journal of Business, Finance & Accounting*, 26(1–2), 103–135.
- Chiappori, P.-A., Levitt, S., & Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, 92, 1138–1151.
- Choi, B., Oppenheimer, D., & Monin, B. (2003, August). *Motivational biases in judgments of streaks in random sequences*. Poster presented at the annual meeting of the European Judgment and Decision Making Society, Zurich, Switzerland.

- Clark, R. D. (2003a). An analysis of streaky performance on the LPGA tour. *Perceptual and Motor Skills*, 97, 365–370.
- Clark, R. D. (2003b). Streakiness among professional golfers: Fact or fiction? *International Journal of Sport Psychology*, 34, 63–79.
- Clark, R. D. (2004). On the independence of golf scores for professional golfers. *Perceptual and Motor Skills*, 98, 675–681.
- Clotfelter, C., & Cook, P. (1993). The gambler's fallacy in lottery play. *Management Science*, 39, 1521–1525.
- Colle, H. A., Rose, R. M., & Taylor, H. A. (1974). Absence of the gambler's fallacy in psychophysical settings. *Perception & Psychophysics*, 15, 31–36.
- Conrad, J., & Kaul, G. (1998). An anatomy of trading strategies. *Review of Financial Studies*, 11, 489–519.
- Cornelius, A., Silva, J. A., III, Conroy, D. E., & Petersen, G. (1997). The projected performance model: Relating cognitive and performance antecedents of psychological momentum. *Perceptual and Motor Skills*, 84, 475–485.
- Crosno, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, 30, 195–209.
- De Bondt, & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40, 793–805.
- De Bondt, & Thaler, R. (1987). Further evidence on investor overreaction and stock-market seasonality. *Journal of Finance*, 42, 557–581.
- Deming, W. E. (1984). *Statistical adjustment of data*. Mineola, NY: Dover.
- Derks, P. L. (1963). Effect of run length on gambler's fallacy. *Journal of Experimental Psychology*, 65, 213–214.
- DiFonzo, N., & Bordia, P. (1997). Rumor and prediction: Making sense (but losing dollars) in the stock market. *Organizational Behavior and Human Decision Processes*, 71, 329–353.
- DiFonzo, N., & Bordia, P. (2002). Rumors and stable-cause attributions in prediction and behavior. *Organizational Behavior and Human Decision Processes*, 88, 785–800.
- Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician*, 58, 38–45.
- Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, 62, 385–394.
- Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). The persistence of risk-adjusted mutual fund performance. *Journal of Business*, 69, 133–157.
- Estes, W. K. (1964). Probability learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 89–123). New York: Academic Press.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83, 37–64.
- Falk, R. (1975). *Perception of randomness*. Unpublished doctoral dissertation, Hebrew University of Jerusalem, Jerusalem, Israel.
- Falk, R. (1981). The perception of randomness. In *Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 222–229). Grenoble, France: Laboratoire IMAG.
- Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104, 301–318.
- Feldman, J. (1959). On the negative recency hypothesis in the prediction of a series of binary symbols. *American Journal of Psychology*, 72, 597–599.
- Feller, W. (1950). *An introduction to probability theory and its applications*. New York: Wiley.
- Fiorina, M. P. (1971). A note on probability matching and rational choice. *Behavioral Science*, 16, 158–166.
- Fisher, I. (1999, January 31). How not to get killed on deadline. *New York Times Magazine*, 36–37.
- Forthofer, R. (1991). Streak shooter—the sequel. *Chance*, 4, 46–48.
- Frame, D., Hughson, E., & Leach, J. C. (2004). *Runs, regimes, rationality: The hot hand strikes back*. Unpublished manuscript, University of Colorado, Leeds School of Business.
- Friedland, N. (1998). Games of luck and games of chance: The effect of luck- versus chance-orientation on gambler's decisions. *Journal of Behavioral Decision Making*, 11, 161–179.
- Frieze, I., & Weiner, B. (1971). Cue utilization and attributional judgments for success and failure. *Journal of Personality*, 39, 591–605.
- Frohlich, C. (1994). Baseball: Pitching no-hitters. *Chance*, 7, 24–30.
- Gawande, A. (1999, February 8). The cancer-cluster myth. *The New Yorker*, 34–37.
- Ghahramani, Z. (2002). An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15, 9–42.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, 103, 592–596.
- Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. *Psychonomic Bulletin and Review*, 2, 260–265.
- Gilovich, T. (1993). *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.
- Gold, E. (1997). *The gambler's fallacy*. Unpublished doctoral dissertation, Carnegie Mellon University, Social and Decision Sciences Department.
- Griffiths, T. L., & Tenenbaum, J. B. (2004). From algorithmic to subjective randomness. *Advances in Neural Information Processing Systems*, 16, 732–737.
- Gronchi, G., & Sloman, S. A. (2008). Do causal beliefs influence the hot-hand and gambler's fallacy? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1164–1168). Austin, TX: Cognitive Science Society.
- Hales, S. (1999). An epistemologist looks at the hot hand in sports. *Journal of the Philosophy of Sport*, 26, 79–87.
- Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, 85, 203–229.
- Hastie, R., & Pennington, N. (2000). Explanation-based decision making. In T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (2nd ed., pp. 212–228). New York: Cambridge University Press.
- Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 344–354.
- Hendricks, D., Patel, J., & Zeckhauser, R. (1993). Hot hand in mutual funds: Short-run persistence of relative performance 1974–1988. *Journal of Finance*, 48, 93–130.
- Huettel, S. A., Mack, P. B., & McCarthy, G. (2002). Perceiving patterns in random series: Dynamic processing of sequence in prefrontal cortex. *Nature Neuroscience*, 5, 485–490.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, 65–91.
- Jakes, S., & Helmsley, D. R. (1986). Individual differences in reaction to brief exposure to unpatterned visual stimulation. *Personality and Individual Differences*, 7, 121–123.
- 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., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kaplan, J. (1990). More on the "hot hand." *Chance*, 3, 6–7.
- Kareev, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1189–1194.
- Keren, G., & Lewis, C. (1994). The two fallacies of gamblers: Type I and

- type II. *Organizational Behavior and Human Decision Processes*, 60, 75–89.
- Keren, G., & Wagenaar, W. A. (1985). On the psychology of playing Blackjack: Normative and descriptive considerations with applications to decision theory. *Journal of Experimental Psychology: General*, 114, 133–158.
- Klaassen, F. J. G. M., & Magnus, J. R. (2001). Are points in tennis independently and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*, 96, 500–509.
- Koehler, J. J., & Conley, C. A. (2003). The hot hand myth in basketball. *Journal of Sport & Exercise Psychology*, 25, 253–259.
- Krantz, D. H., Atkinson, R. C., Luce, R. D., & Suppes, P. (Eds.). (1974). *Contemporary developments in mathematical psychology: Learning memory, and thinking* (Vol. 1). San Francisco: Freeman.
- Ladouceur, R., Paquet, C., & Dubé, D. (1996). Erroneous perceptions in generating sequences of random events. *Journal of Applied Social Psychology*, 26, 2157–2166.
- Laplace, P. (1995). *Philosophical essay on probabilities* (A. Dale, Trans.). New York: Springer Verlag. (Original work published 1825).
- Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the "hot hand." *Chance*, 2, 22–30.
- Lindman, H., & Edwards, W. (1961). Supplementary report: Unlearning the gambler's fallacy. *Journal of Experimental Psychology*, 62, 630.
- Lock, R. (2003). A streak like no other. *Statistics*, 38, 24–26.
- Lopes, L. L. (1982). Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 626–636.
- Lopes, L. L., & Oden, G. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 13, 392–400.
- Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121.
- Marsaglia, G., Zaman, A., & Tsang, W. W. (1990). Towards a universal random number generator. *Statistics & Probability Letters*, 8, 35–39.
- Matthews, L., & Sanders, W. (1984). Effects of causal and non-causal sequences of information on subjective prediction. *Psychological Reports*, 54, 211–215.
- McClelland, G. H., & Hackenberg, B. H. (1978). Subjective probabilities for sex of next child: U.S. college students and Philippine villagers. *Journal of Population*, 1, 132–147.
- Metrick, A. (1999). Performance evaluation with transactions data: The stock selection of investment newsletters. *Journal of Finance*, 54, 1743–1775.
- Metzger, M. A. (1985). Biases in betting: An application of laboratory findings. *Psychological Reports*, 56, 883–888.
- Miller, S., & Weinberg, R. (1991). Perceptions of psychological momentum and their relationship to performance. *The Sports Psychologist*, 5, 211–222.
- Miyoshi, H. (2000). Is the "hot hands" phenomenon a misperception of random events. *Japanese Psychological Research*, 42, 128–133.
- Moldoveanu, M., & Langer, E. (2002). False memories of the future: A critique of the applications of probabilistic reasoning to the study of cognitive processes. *Psychological Review*, 109, 358–375.
- Morris, M. W., Sheldon, O. J., Ames, D. R., & Young, M. J. (2005). *Metaphor in stock market commentary: Consequences and preconditions of agentic descriptions of price trends*. Unpublished manuscript, Columbia University.
- Morrison, D. G., & Schmittlein, D. C. (1998). It takes a hot goalie to raise the Stanley Cup. *Chance*, 11, 3–7.
- Morrison, R. S., & Ordeshook, P. C. (1975). Rational choice, light guessing and the gambler's fallacy. *Public Choice*, 22, 79–89.
- Morse, E. E., & Rundquist, W. N. (1960). Probability matching with an unscheduled random sequence. *American Journal of Psychology*, 73, 603–607.
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *Journal of Finance*, 54, 1249–1290.
- Mullainathan, S. (2002). *Thinking through categories*. Unpublished manuscript, Harvard University, Department of Economics.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Neuringer, A. (1986). Can people behave randomly? The role of feedback. *Journal of Experimental Psychology: General*, 115, 62–75.
- Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, 109, 330–357.
- Nicks, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *Journal of Experimental Psychology*, 57, 105–114.
- Odean, T. (1998). Are investors reluctant to realize their losses? *Journal of Finance*, 53, 1775–1789.
- Offerman, T., & Sonnemans, J. (2004). What's causing overreaction? An experimental investigation of recency and the hot-hand effect. *Scandinavian Journal of Economics*, 106, 533–553.
- Olivola, C. Y., & Oppenheimer, D. M. (2008). Randomness in retrospect: Exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review*, 15, 991–996.
- O'Neill, B. (1987). Nonmetric test of the minimax theory of two-person zero-sum games. *Proceedings of the National Academy of Sciences, USA*, 84, 2106–2109.
- Palacios-Huerta, I. (2003). Professionals play minimax. *Review of Economic Studies*, 70, 395–415.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Poincaré, H. (1952). *Science and method* (Francis Maitland, Trans.). London: Dover. (Original work published 1914).
- Rabin, M. (2002, August). Inference by believers in the law of small numbers. *Quarterly Journal of Economics*, 775–816.
- Rabiner, L. R., & Juang, B. H. (1986, February). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3, 4–16.
- Rapoport, A., & Budescu, D. (1992). Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General*, 121, 352–363.
- Rapoport, A., & Budescu, D. (1997). Randomization in individual choice behavior. *Psychological Review*, 104, 603–617.
- Reichenbach, H. (1949). *The theory of probability* (E. Hutten & M. Reichenbach, Trans.). Berkeley: University of California Press. (Original work published 1934).
- Restle, F. (1961). *Psychology of judgment and choice: A theoretical essay*. New York: Wiley.
- Richardson, P. A., Adler, W., & Hanks, D. (1988). Game, set, match: Psychological momentum in tennis. *The Sport Psychologist*, 2, 60–76.
- Rodgers, J. L., & Doughty, D. (2001). Does having boys or girls run in the family? *Chance*, 14, 8–13.
- Roney, C. J. R., & Trick, L. M. (2003). Grouping and gambling: A gestalt approach to understanding the gambler's fallacy. *Canadian Journal of Experimental Psychology*, 57, 69–75.
- Schank, R. C., Collins, G. C., & Hunter, L. E. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences*, 9, 639–651.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15, 233–250.
- Shaw, J. M., Dziewaltoski, D. A., & McElroy, M. (1992). Self-efficacy and causal attribution as mediators of perceptions of psychological momentum. *Journal of Sport & Exercise Psychology*, 14, 134–147.
- Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losses too long: Theory and evidence. *Journal of Finance*, 40, 777–790.
- Silva, J. M., Hardy, C. J., & Crace, R. K. (1988). Analysis of psychological

- momentum in intercollegiate tennis. *Journal of Sport & Exercise Psychology*, 10, 346–354.
- Sirri, E., & Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*, 53, 1589–1622.
- Smith, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin and Review*, 10, 753–758.
- Stern, H. S. (1997). Judging who's hot and who's not. *Chance*, 10, 40–43.
- Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, 1, 1–12.
- Teigen, K. H. (2005). When a small difference makes a big difference: Counterfactual thinking and luck. In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 129–146). New York: Routledge.
- Terrell, D. (1994). A test of the gambler's fallacy—Evidence from pari-mutuel games. *Journal of Risk and Uncertainty*, 8, 309–317.
- Terrell, D. (1998). Biases in assessments of probabilities: New evidence from Greyhound races. *Journal of Risk and Uncertainty*, 17, 151–166.
- 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.
- Tune, G. S. (1964). Response preferences: A review of some relevant literature. *Psychology Bulletin*, 61, 286–302.
- Tversky, A., & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance*, 2, 16–21.
- Tversky, A., & Gilovich, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance*, 2, 31–37.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110.
- Tyszka, T., Zielonka, P., Dacey, R., & Sawicki, P. (2008). Perception of randomness and predicting uncertain events. *Thinking & Deciding*, 14, 83–110.
- Vergin, R. (2000). Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior*, 23, 181–197.
- Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 14, 101–118.
- Wagenaar, W. A. (1970). Appreciation of conditional probabilities in binary sequences. *Acta Psychologica*, 34, 348–356.
- Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of the literature. *Psychological Bulletin*, 77, 65–72.
- Wagenaar, W. A. (1988). *Paradoxes of gambling behavior*. London: Erlbaum.
- Wagenaar, W. A. (1991). Randomness and randomizers: Maybe the problem is not too big. *Journal of Behavioral Decision Making*, 4, 220–222.
- Wagenaar, W. A., & Keren, G. (1988). Chance and luck are not the same. *Journal of Behavioral Decision Making*, 1, 65–75.
- Walker, M., & Wooders, J. (2001). Minimax play at Wimbledon. *American Economic Review*, 91, 1521–1538.
- Wardrop, R. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, 49, 24–28.
- Wilke, A., & Barrett, H. C. (in press). The hot hand phenomenon as a cognitive adaptation for clumped resources. *Evolution and Human Behavior*.
- Yackulic, R. A., & Kelly, I. W. (1984). The psychology of the "gambler's fallacy" in probabilistic reasoning. *Psychology—A Quarterly Journal of Human Behavior*, 21, 55–58.

Received November 20, 2007

Revision received October 20, 2008

Accepted October 24, 2008 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!