Title
Detecting the Hot Hand: An Alternative Model

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Author
Sun, Yanlong

Publication Date
2004

Peer reviewed

Detecting the Hot Hand: An Alternative Model

Yanlong Sun (Yanlong.Sun@uth.tmc.edu)

University of Texas Health Science Center at Houston
School of Health Information Sciences, 7000 Fannin Suite 600
Houston, TX 77030 USA

Abstract

The belief in the hot hand was suggested to be a "cognitive illusion" since no significant evidence was found in the basketball-shooting data to reject the simple binomial model (Gilovich, Vallone & Tversky, 1985). The present study argues that in order to evaluate the validity of human perception and cognition such as the hot hand belief, a data-driven approach is needed to compare multiple alternative models. A hot hand model with nonstationary shooting accuracy was tested and showed significantly better approximation to the data than the binomial model, indicating that the simple binomial model may not be accurate enough to serve as a normative model. This finding suggests that the hot hand might indeed have existed, and weakens the argument that the hot hand belief might be "seeing patterns out of randomness."

The Hot Hand and the Perception of Randomness

The "hot hand" in the game of basketball has received much attention in cognitive psychology because it touches an interesting topic about human perception and cognition of random and non-random events outside the psychological laboratory. A long-lasting debate about whether the hot hand exists, hence, whether the hot hand belief is a valid cognitive activity, was triggered by three articles by Gilovich, Vallone and Tversky (1985), and Tversky and Gilovich (1989a, 1989b) (later "GVT" refers to these three articles as a group, unless specified otherwise). The researchers interpreted the hot hand belief as a manifestation about statistically significant deviations from what is expected by the simple binomial model, namely, nonstationary shooting accuracy or positive dependence in basketball shooting sequences. However, no statistical evidence was found to support such belief. After a number of statistical analyses on a large set of data, the researchers found that actual basketball shooting sequences were "indistinguishable from that produced by a simple binomial model" (Gilovich et al., 1985, p. 297). They concluded, "perhaps, then, the belief in the hot hand is *merely* [italics added] one manifestation of this fundamental misconception of the laws of chance" (Tversky & Gilovich, 1989a, p. 16).

Since GVT, many studies have been carried out to investigate the hot hand in basketball or other sports such as baseball. These studies roughly fell into four categories: a) studies that conducted null hypothesis tests but failed to reject the binomial model (e.g., Adams, 1992; Albright, 1993; Chatterjee, Yilmaz, Habibullah, & Laudato, 2000), (b) studies that raised concerns about the power of significance tests conducted by Gilovich et al. (1985) and Albright (1993) (e.g., Miyoshi, 2000; Stern & Morris, 1993; Sun, 2001, 2003; Wardrop, 1999), (c) studies that proposed alternative models that may support the hot hand belief (e.g., Albert, 1993; Albert & Bennett, 2001; Larkey, Smith, & Kadane, 1989), (d) a study that addressed the adaptive value of the hot hand belief, assuming the accuracy of the binomial model (Burns, 2001).

The present paper takes a step further and examines the accuracy of the simple binomial model in a side-by-side comparison with an alternative model that assumes the existence of the hot hand. The importance of such a comparison is obvious since which model is more accurate would inevitably affect researchers' opinion about the validity of the hot hand belief. As Brunswik (1956) and Simon (1982) suggested, the environment in which human perception and cognition originate and operate must be carefully studied. On one hand, it is possible that the hot hand does not exist and the hot hand belief is another example of misperceptions of randomness outside the psychological laboratory, in addition to many previous findings when random events were clearly defined (e.g., Falk, 1981; Kahneman and Tversky, 1972; Tversky & Kahneman, 1971, 1974; Wagenaar, 1972). On the other hand, it is possible that the hot hand does exist, even with a substantial effect size (e.g., substantial changes in shooting accuracy), and traditional statistical tests are generally low in power thus not capable of detecting the effect. The fact is that a truly random process can produce seemingly non-random "patterns," but a truly non-random process can produce seemingly random events as well. Lopes and Oden (1987) demonstrated that although human subjects sometimes misidentified random events as nonrandom (i.e., false alarms), they could also correctly detect truly nonrandom signals (i.e., correct hits). Thus, it is important to find out whether the simple binomial model is accurate enough to serve as a normative model. Then, researchers might be able to answer the question whether the hot hand belief is more about signal detections, or, just "seeing something out of randomness."

Model-driven vs. Data-driven

GVT concluded that actual basketball-shooting records "may be *adequately* [italics added] described by a simple binomial model" (Gilovich et al., 1985, p. 313). However, such a conclusion was solely based on the non-significant *p* values in null hypothesis tests under the binomial model. Sun (2003) and Wardrop (1999) pointed out that GVT's statistical tests were largely redundant and generally low in power, and in many cases, GVT failed to report large deviations from the binomial process or misinterpreted the test results. In the present paper, I only address the importance of comparing multiple models and why non-significant *p* values do not necessarily suggest the accuracy of the simple binomial model.

Criticisms of null hypothesis significance testing (NHST) have been leveled for decades. Many researchers warned that when alternative hypotheses abound, misinterpretations of statistical significance could easily arise (e.g., Cohen, 1994; Lykken, 1991; Oakes, 1986). Nevertheless, many researchers tend to ignore the fact that NHST only estimates $p(D \mid H_0)$, the probability that data *D* could have arisen if the null hypothesis H_0 were true, not $p(H_0 \mid D)$, the probability that H_0 is true, given *D*. In modeling basketball shooting, the fact that no significant deviation was found to reject the binomial model, namely, $p(D \mid H_{Binomial}) > .05$, only indicates that the binomial model may not be terribly erroneous. However, not being terribly erroneous is not the same thing as being accurate or being unique. A *p* value greater than .05 only prompts researchers to retain the null, not to accept the null as if it were true or even likely to be true.

Let $H_{Binomial}$ denote the event that the binomial model is true, $H_{Hot\ Hand}$ denote the event that the hot hand theory is true, and *D* denote the event that a certain statistic from the shooting data reaches a certain level. In order to demonstrate the adequacy of binomial model or the invalidity of the hot hand theory, given the available data, one needs to find out which hypothesis the data are in favor of, namely, to compare $p(H_{Binomial} \mid D)$ and $p(H_{Hot\ Hand} \mid D)$. In Bayes' theorem,

$$\frac{p(H_{Binomial} \mid D)}{p(H_{Hot\ Hand} \mid D)} = \frac{p(H_{Binomial})\, p(D \mid H_{Binomial})}{p(H_{Hot\ Hand})\, p(D \mid H_{Hot\ Hand})}. \quad (1)$$

If one is not biased toward either one of the two hypotheses before examining the data, it is reasonable to assign equal prior probabilities to both models, $p(H_{Binomial}) = p(H_{Hot\ Hand}) = .50$. Then, the comparison between $p(H_{Binomial} \mid D)$ and $p(H_{Hot\ Hand} \mid D)$ comes down to the comparison between $p(D \mid H_{Binomial})$ and $p(D \mid H_{Hot\ Hand})$. GVT's statistical analyses showed that in a number of statistical tests,[1] $p(D \mid H_{Binomial})$ was not significantly small. Nevertheless, such information alone cannot invalidate the hot hand theory, another piece of information, $p(D \mid H_{Hot\ Hand})$ is still missing.

The argument here actually calls for a data-driven approach that compares at least two rival models, rather than a model-driven approach that conducts null hypothesis tests only on one model. The distinction between these two approaches is not a clear cut but rather a difference in emphasis. The data-driven approach eventually has to come down to evaluations of a limited number of models one by one. If a certain model superior to others arises, it will be tested against further data for a need to abandon or modify the model. In this sense, the distinction between two rival models often is not an absolute dichotomy. It is true that in hypothesis testing, such as in Equation 1, two hypotheses have to be exclusive to each other. Nevertheless, in data modeling, two models might only differ in the degrees they approximate the actual process. Which model is selected would be based on which model provides a better approximation of the data, rather than some "mechanical dichotomous decisions around a sacred .05 criterion" (Cohen, 1994, p. 997).

Extracting Relevant Statistics from the Data

To compare multiple models by a data-driven approach, it is essential to extract relevant statistics from the available data. Sun (2003) pointed out that the statistical tests conducted by GVT, such as the test of serial correlation (compared to zero) and runs test were largely focused on the first moment estimate of the time series, namely, the *hit rate* (i.e., observed hitting percentage in a sequence of a certain length) as an estimate of *shooting accuracy* (i.e., the probability for any given shot to be a hit). However, by the law of large numbers, *hit rate* only provides a good approximation of *shooting accuracy* when shooting accuracy remains constant and the sample size is considerably large. Thus, assuming the hot hand is about the nonstationarity of the shooting accuracy, fluctuations of shooting accuracy would not be easily detected by fluctuations of hit rate, when a player only took a limited number of shots in each game. For instance, given a result of 5 hits in a sequence of 10 shots, a null hypothesis test *alone* cannot distinguish whether the hit rate of 50% is a result of a shooting accuracy of 40% or a shooting accuracy of 60%.

By focusing on higher moments of the shooting sequences, Sun (2003) found significant fluctuations of serial correlations in the field goal data that were originally reported by Gilovich et al. (1985). That is, a player sometimes shot in streaks (i.e., successive hits or misses), such as in {1, 1, 1, 1, 0, 0, 0, 0}, yielding a positive serial correlation, and sometimes shot alternatively (hits and misses alternated very often), such as in {1, 0, 1, 0, 1, 0, 1, 0}, yielding a negative serial correlation. The observed changes in serial correlations were unlikely to be accounted for by the simple binomial model, namely, $p(D \mid H_{Binomial}) < .05$, where *D* represents the event that the serial correlations changed significantly. Only when the data were *aggregated* across all periods, the *overall averaged* serial correlation was close to zero (e.g., comparing the overall

[1] Note that most of GVT's tests were mathematically redundant (see Wardrop, 1999).

serial correlation with zero, $p > .05$). This finding has at least two indications. First, the actual basketball shooting might not be a stationary process since hits and misses are *not evenly* distributed in the observed shooting sequence. Second, fluctuations of hit rates and the overall serial correlation are not sensitive enough to capture such nonstationarity. In the following, I will present an alternative model that can be distinguished from the simple binomial model by examining the fluctuations of serial correlations. Furthermore, this model may provide a better approximation to the observed data.

A Model of the Hot Hand

Model and Parameter Settings

In real basketball games, it is very possible that potentially high or low shooting accuracy ("hot hand" or "cold hand") might exist but were *interrupted* by other activities such as shot selection and defensive pressure. For example, after making one or two shots, a player may become confident and try more difficult shots, or the opposing team may intensify their defensive pressure on that player. Less frequent interruptions tend to produce shooting sequences with positive serial correlations, since the player's shooting accuracy, either high or low, remains comparatively unchanged, for example, an extreme case would be something like {1, 1, 1, 1, 1, 0, 0, 0, 0, 0}. And vice versa, more frequent interruptions tend to produce shooting sequences with negative serial correlations, for example, a resulting sequence like {1, 0, 1, 0, 1, 0, 1, 0, 1, 0}.

Figure 1 represents a Markov switching model (hence referred to as "the hot hand model"). Similar models have been used by Lopes and Oden (1987) in studying human subjects' ability of distinguishing between random and nonrandom events, and by Albert and Bennett (2001) in modeling the "streakiness" in baseball.

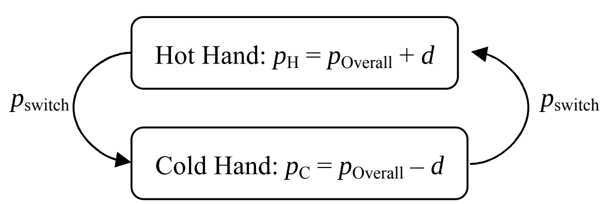

Figure 1. A Markov Model of the Hot Hand

To accommodate the hot hand theory, the major characteristic of this model is that it has two states, "hot hand" and "cold hand," representing two different levels of shooting accuracies, p_H and p_C, respectively. If a player's overall shooting percentage in the entire season was $p_{Overall}$, p_H and p_C were shifted higher or lower in the same amount of d from $p_{Overall}$. Then, this player's simulated shooting sequence will be generated as the player switches between the "hot hand" and the "cold hand." How often the player makes the switch depends on the switching probability, p_{switch}. A high p_{switch} value (e.g., $p_{switch} > .50$) means the player switches between two states very often. In an actual basketball game, this would represent the situation in which a hot hand or a cold hand is detected and a real-time adjustment is immediately deployed by either the player or the opposing team. And vice versa, a low p_{switch} value (e.g., $p_{switch} < .50$) means that the player rarely switches between two states. This would represent the situation in which a hot hand or a cold hand remained uninterrupted or real-time adjustments rarely occurred.

Actually, when $p_H = p_C = p_{Overall}$ ($d = 0$) and $p_{switch} = .50$, the hot hand model is in effect equivalent to the binomial model. If the binomial model were truly adequate and unique, one would expect that a model with dramatically different parameter settings would be less capable of describing the observed data. For this reason, I chose a set of extreme values to represent the hot hand model, in which $d = .30$ (i.e., $p_H - p_C = .60$) and p_{switch} was randomly selected from (.95 and .05) with a 50-50 percent chance for every 10 shots, whereas the binomial model only took a constant shooting accuracy $p_{Overall}$. Figure 2 illustrates the difference between two models in terms of the shooting accuracy along the time line.

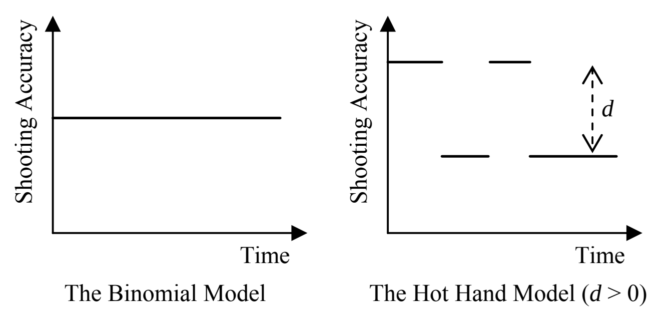

Figure 2. Two Possible Models of Basketball Shooting

Simulation Procedure

Gilovich has kindly provided the field goal data that were reported in Gilovich et al. (1985). There were 18 players in the data set, and 16 of them were included in the simulation (2 players were excluded because their shooting sequences were too short).

For each player in the simulation, I computed a statistic called "MMAC" (Max-Min Moving Autocorrelation) from his actual shooting sequence, whereas MMAC was defined as the absolute difference between the largest and smallest moving serial correlations, where the moving serial correlations were calculated as the serial correlations within a window of 100 shots, starting from the first shot then each time moving 1 shot further until the end of the sequence. The purpose for choosing such specific statistic is to capture the fluctuations of the serial correlations. In the meantime, to reduce chance errors, a large sample size is needed so that the window width of 100 shots was chosen.

For each of the 16 players, I ran 10,000 simulations with the binomial model and another 10,000 simulations with the hot hand model, each simulation generating one shooting sequence in the same length of the player's actual shooting sequence and with the same overall shooting accuracy. The statistic MMAC was calculated from each simulated sequence, then compared to the observed MMAC from the player's actual shooting record. The probabilities for each model's simulated MMAC to include the observed MMAC were computed as $p(D \mid H_{\text{Binomial}})$ and $p(D \mid H_{\text{Hot Hand}})$. Then, given equal prior probabilities $p(H_{\text{Binomial}}) = p(H_{\text{Hot Hand}}) = .50$, posterior probabilities $p(H_{\text{Binomial}} \mid D)$ and $p(H_{\text{Hot Hand}} \mid D)$ were calculated by Equation 1. Since there were only two hypotheses considered, $p(H_{\text{Binomial}} \mid D) + p(H_{\text{Hot Hand}} \mid D) = 1$.

Simulation Results

The simulation results are listed in Table 1. Columns 2 to 5 list the probabilities $p(D \mid H_{\text{Binomial}})$, $p(D \mid H_{\text{Hot Hand}})$, $p(H_{\text{Binomial}} \mid D)$, and $p(H_{\text{Hot Hand}} \mid D)$, respectively. Column 6 lists the probabilities of detecting significance (a = .05, two-tailed) by runs test (Siegel, 1956) on the sequences generated by the hot hand model. The table is ordered in the ascending order of $p(D \mid H_{\text{Binomial}})$.

Considered separately, the probabilities $p(D \mid H_{\text{Binomial}})$ and $p(D \mid H_{\text{Hot Hand}})$ (Columns 2 and 3) in effect provided p values for null hypothesis significance testing, assuming either of the two models as the ture hypothesis (a = .05, two-tailed). For players 24, 10, and 3, the simulation results $p(D \mid H_{\text{Binomial}}) < .05$ actually provided significant p values to reject the binomial model. For players 18 and 50, $p(D \mid H_{\text{Binomial}})$ were only slightly greater than .05. (Considering the fact that there were 16 players tested, the probability of family-wise Type I errors needs to be calculated, which was found to be less than .05. see Sun, 2003) On the other hand, none of the p values in $p(D \mid H_{\text{Hot Hand}})$ reached the significance level of .05.

Assuming one is unbiased toward either of the two models prior to examining the data, so that $p(H_{\text{Binomial}}) = p(H_{\text{Hot Hand}}) = .50$, the comparisons between $p(H_{\text{Binomial}} \mid D)$ and $p(H_{\text{Hot Hand}} \mid D)$ (Columns 4 and 5) would reveal which model obtains more support from the observed data in terms of the MMAC statistic.

Table 1. Comparisons between the binomial model and the hot hand model

Player	$p\ (D \mid H_{\text{Binomial}})$	$p\ (D \mid H_{\text{Hot Hand}})$	$p\ (H_{\text{Binomial}} \mid D)$	$p\ (H_{\text{Hot Hand}} \mid D)$	Power (runs test)
24	.0178	.3380	.0500	.9500	.1911
10	.0223	.3127	.0666	.9334	.1909
3	.0232	.4116	.0534	.9466	.1891
18	.0508	.1299	.2811	.7189	.1836
50	.0690	.4709	.1278	.8722	.1824
7	.1517	.5929	.2037	.7963	.1854
25	.4084	.6539	.3844	.6156	.1836
2	.5343	.9610	.3573	.6427	.1951
11	.5446	.8983	.3774	.6226	.1795
22	.6472	.9640	.4017	.5983	.1769
53	.7094	.9766	.4208	.5792	.1936
5	.7370	.9855	.4279	.5721	.1928
4	.7625	.9953	.4338	.5662	.1918
6	.8004	.9993	.4447	.5553	.1872
1	.9393	.9993	.4845	.5155	.1871
9	.9886	.9999	.4972	.5028	.1845
Mean	.4632	.7297	.3161	.6839	.1872

Note: D represents the event that the simulated MMAC is greater than or equal to the observed MMAC calculated from each player's shooting record. Column 6 is the estimated power of runs test based on detections of significance (a = .05, two-tailed) on the simulated sequences by the hot hand model.

For individual cases, MMAC appeared to be substantially in favor of the hot hand model rather than the binomial model for a certain number of players (e.g., players 24, 10, 3, 18, 50, 7, 25, 2, and 11), as $p(H_{\text{Binomial}} \mid D)$ was much smaller than $p(H_{\text{Hot Hand}} \mid D)$ (see Table 1, Columns 4 and 5). One may calculate a c^2 statistic for each player to test the null hypothesis that MMAC is indifferent to either of the binomial model or the hot hand model. However, c^2 statistics tend to be over-sensitive when the expected frequency in a certain cell is too low (for example, the players 6, 1, and 9). The result that all c^2 were significant ($df = 1$, $p < .01$) for all of the 16 players might have overestimated the superiority of the hot hand model.

Taking all 16 players together, the hot hand model appeared to be substantially superior to the binomial model in accounting for the observed MMAC. On the average, $p(D \mid H_{\text{Binomial}}) = .4632$, and $p(D \mid H_{\text{Hot Hand}}) = .7297$. By the criterion of maximum likelihood, given equal priors $p(H_{\text{Binomial}}) = p(H_{\text{Hot Hand}}) = .50$, the observed data seemed to support the hot hand model rather than the binomial model: on the average, the posterior probabilities are $p(H_{\text{Binomial}} \mid D) = .3161$ and $p(H_{\text{Hot Hand}} \mid D) = .6839$.

It may be possible that the hot hand model appeared to be superior to the binomial model only in terms of the statistics of MMAC. To see whether the hot hand model was "truthful" to other observed statistics such as the number of runs, I also conducted a runs test for each simulated sequence by the hot hand model, since out of those 16 players, runs test only detected one significance at the .05 level in the observed shooting sequence (player 53, see Gilovich et al., 1985). (Note that because of the symmetrical setting of the model, there is no need to check the hitting percentage.) The results of runs test suggested that the hot hand model was largely truthful to the observed shooting sequence in the statistic of number of runs, since on the average, only 18.72% of the simulated sequences were detected as significant deviations from what is expected by the binomial model (see Table 4, last column). A further check found that during 10,000 simulations for each player with the hot hand model, the *overall* serial correlations were symmetrically distributed around the mean of zero, with a standard deviation slightly larger than the expected value ($1/\sqrt{N-3}$) assuming binomial process (N is the number of shots in each sequence). Together, these observations provided confirmations to my previous claims. That is, a nonrandom process (such as the hot hand model) can produce seemingly random sequences and may not be easily detected by traditional statistical methods (such as the runs test, or, comparing the overall serial correlation with zero).

Discussion

One might argue that the "hot hand model" fitted the data better than the binomial model simply because the former has more parameters than the latter. I have three reasons to counter this argument. First, basketball shooting is a complex process. It is very reasonable to believe that a useful model needs more parameters than just a single constant shooting accuracy. Second, the extra parameters in the hot hand model may not be counted as "free parameters" because they feasibly represent actual situations in which a player's shooting accuracy may change and real-time adjustments take place quickly (or slowly). Lastly and most importantly, as mentioned before, the hot hand model actually took parameter values that were substantially different from the simple binomial model. Yet, it provided more accurate descriptions of the observed data. This would have seriously challenged the accuracy of the simple binomial model.

It should be pointed out that the primary purpose for building the hot hand model is not to argue about its uniqueness. Nevertheless, such model may prompt researchers to consider the possibility that non-random process may easily produce seemingly random sequences and the possibility that the hot hand belief is indeed a valid cognitive activity in detecting non-random events. It is important to notice that particular statistics such as number of runs, serial correlations, including the MMAC statistic I used in this study, may not be sensitive enough to tell the difference between two different processes. Nevertheless, researchers need to consider multiple models in evaluating the validity of human perceptions, since multiple models can co-exist and provide different levels of approximations to the actual underlying process.

The simulation has shown that for a certain number of players, the hot hand model is substantially superior to the binomial model. For the other players, these two models are not easily distinguishable. By Bayes' theorem in Equation 1, if both models account for the data with the same capability so that $p(D \mid H_{\text{Binomial}}) \approx p(D \mid H_{\text{Hot Hand}})$, which model is more likely to be "perceived" from the data, namely, $p(H_{\text{Binomial}} \mid D)$ and $p(H_{\text{Hot Hand}} \mid D)$, then, is entirely determined by personal beliefs, $p(H_{\text{Binomial}})$ and $p(H_{\text{Hot Hand}})$. There is no prior reason why basketball fans and players should agree with researchers on such personal belief. In other words, the hot hand belief may not be readily dismissed as merely a misperception of randomness simply because the researchers failed to reject the binomial model by null hypothesis significance testing.

General Conclusion

The primary purpose of the present paper is not to dispute whether ordinary people misperceive probabilistic events in basketball, but to prompt further investigations of the actual process of basketball shooting. Lacking normative knowledge such as probability theory and theories of stochastic processes, ordinary people are often prone to mistakes. However, it is also possible that the hot hand belief was describing a true anomaly that was not detected by traditional statistical methods. The present study presented a case when statistical methods are applied objectively rather than subjectively toward the plausible models, how a different point of view, regarding the validity of human perceptions of the environment, could be obtained. That is, comparing to a model-driven approach that only conducts null hypothesis testing on a single model, a data-driven approach can be more revealing by comparing multiple models. Then, it was suggested that the simple binomial model might not be accurate enough to serve as a normative model in evaluating the validity of the hot hand belief. From Brunswik's (1956) point of view, an organism and the environment in which the organism was embedded should receive equal emphasis in psychological theory and research. In this sense, the primary purpose of the present study is to serve as "a propaedeutic to functional psychology" (Brunswik, 1956, p. 119), a necessary step before psychologists can fully understand the belief in the hot hand.

Acknowledgements

This research was supported in part by the postdoctoral fellowship from the W. M. Keck Center for Computational and Structural Biology of the Gulf Coast Consortia.

References

Adams, R.M. (1992). The "hot hand" revisited: Successful basketball shooting as a function of intershot interval. *Perceptual and Motor Skills, 74*, 934.

Albert, J. (1993). A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association, 88(424)*, 1184-1188.

Albert, J., & Bennett, J. (2001). *Curve ball: Baseball, statistics, and the role of chance in the game*. New York: Springer-Verlag.

Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association, 88(424)*, 1175-1183.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley, CA: University of California Press.

Burns, B. D. (2001). The hot hand in basketball: Fallacy or adaptive thinking? In J. D. Moore, & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society* (pp. 152-157). Hillsdale, NJ: Lawrence Erlbaum.

Chatterjee, S., Yilmaz, M. R., Habibullah, M., & Laudato, M. (2000). An approximate entropy test for randomness. *Communications in Statistics: Theory and Methods, 29(3)*, 655-675.

Cohen, J. (1994). The earth is round (p<0.05). *American Psychologist, 49*, 997-1003.

Falk, R. (1981). The perception of randomness. In C. Comiti, & G. Vergnaud (Eds.), *Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (pp. 222-229). Grenoble, France.

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

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

Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's Okay to believe in the "hot hand." *Chance 2(4)*, 22-30.

Lopes, L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13(3)*, 392-400.

Lykken, D. L. (1991). What's wrong with psychology? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology, vol. 1: Matters of public interest, essays in honor of Paul E. Meehl* (pp. 3-39). Minneapolis, MN: University of Minnesota Press.

Miyoshi, H. (2000). Is the "hot hands" phenomenon a misperception of random events? *Japanese Psychological Research, 42(2)*, 128-133.

Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw-Hill.

Simon, H. A. (1982). *Models of bounded rationality*, Vol.3, Cambridge, MA: MIT Press.

Stern. H. S., & Morris, C. N. (1993). A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association, 88(424)*, 1189-1194.

Sun, Y. (2003). *The hot hand revisited: Toward a data-driven approach.* Manuscript submitted for publication.

Sun, Y. & Tweney, R. D. (2001, November). *Detecting the "hot hand": A time series analysis of basketball*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

Tversky, A., & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance, 2(1)*, 16-21.

Tversky, A., & Gilovich, T. (1989b). The "hot hand": Statistical reality or cognitive illusion? *Chance, 2(4)*, 31-34.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

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

Wardrop, R. L. (1999). *Statistical tests for the hot hand in basketball in a controlled setting* (Tech. Rep.). University of Wisconsin-Madison, Department of Statistics.