Ecological Psychology

ISSN: 1040-7413 (Print) 1532-6969 (Online) Journal homepage: http://www.tandfonline.com/loi/heco20

Toward an Ecological Theory of Rationality: Debunking the Hot Hand "Illusion"

Taleri Hammack, Jehengar Cooper, John M. Flach & Joseph Houpt

To cite this article: Taleri Hammack, Jehengar Cooper, John M. Flach & Joseph Houpt (2017) Toward an Ecological Theory of Rationality: Debunking the Hot Hand "Illusion", Ecological Psychology, 29:1, 35-53, DOI: 10.1080/10407413.2017.1270149

To link to this article: http://dx.doi.org/10.1080/10407413.2017.1270149

Published online: 31 Jan 2017.

ECOLOGICAL PSYCHOLOGY
2017, VOL. 29, NO. 1, 35–53
http://dx.doi.org/10.1080/10407413.2017.1270149

Routledge
Taylor & Francis Group

Toward an Ecological Theory of Rationality: Debunking the Hot Hand "Illusion"

Taleri Hammack, Jehengar Cooper*, John M. Flach, and Joseph Houpt

Department of Psychology, Wright State University

ABSTRACT

This article explores the "hot hand illusion" from the perspective of ecological rationality. Monte Carlo simulations were used to test the sensitivity of typical tests for randomness to plausible constraints (e.g., Wald-Wolfowitz) on sequences of binary events (e.g., basketball shots). Most of the constraints were detected when sample sizes were large. However, when the range of improvement was limited to reflect natural performance bounds, these tests did not detect a success-dependent learning process. In addition, a series of experiments assessed people's ability to discriminate between random and constrained sequences of binary events. The result showed that in all cases human performance was better than chance, even for the constraints that were missed by the standard tests. The case is made that, as with perception, it is important to ground research on human cognition in the demands of adaptively responding to ecological constraints. In this context, it is suggested that a "bias" or "default" that assumes that nature is "structured" or "constrained" is a very rational approach for an adaptive system whose survival depends on assembling smart mechanisms to solve complex problems.

> Illusions will be treated as special cases of perception, not as phenomena which might reveal the laws of the subjective process of perception. (Gibson, 1963/1982, p. 366)

> I am asking the reader to suppose that the concept of space has nothing to do with perception. Geometrical space is a pure abstraction. Outer space can be visualized but it cannot be seen. The cues for depth refer only to paintings, nothing more. …
> The doctrine that we could not perceive the world around us unless we already had the concept of space is nonsense. It is quite the other way around: We could not conceive of empty space unless we could see the ground under our feet and the sky above. Space is a myth, a ghost, a fiction for geometers. (Gibson, 1979/1986, p. 3)

A goal of our current research program is to broaden the insights from the ecological approach to perception to address broader aspects of cognition (e.g., decision making and problem solving; Flach & Voorhorst, 2016). A motivating assumption of this research is to examine human cognition in the context of a closed-loop, functional coupling between an

CONTACT John M. Flach john.flach@wright.edu Department of Psychology, Wright State University, 3640 Colonel Glenn Highway, Dayton, OH 45435.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/heco.

*Jehengar Cooper is currently Human Performance Analyst at Canadian Nuclear Laboratories.

actor and an ecology. We believe such an approach is consistent with the spirit of Gibson and earlier functionalist psychologists/philosophers (e.g., Dewey, James, Peirce) who viewed cognition as a capacity for adapting to the pragmatic demands of survival in a complex world. Today, this spirit is reflected in the work of Gigerenzer and colleagues in their construct of Ecological Rationality (e.g., Todd, Gigerenzer, & The ABC Research Group, 2012).

To make the link to the Gibsonian approach to perception explicit, we make the following claims about research on decision making and problem solving that are analogous to Gibson's claims quoted at the beginning of this article (italics indicate changes to Gibson's claims to reflect an extension of the ecological framework to cognition/decision making/problem solving):

Claim 1. *Biases or heuristics* are treated as special cases of *cognition*, not as phenomena that might reveal the laws of the subjective process of *rationality*.

Claim 2. *We* are asking the reader to suppose that the concept of *logic* has nothing to do with *problem solving*. *Absolute truth or validity* is a pure abstraction. *Absolute truth* can be *imagined* but it cannot be *realized*. The *logical norms for validity* refer only to *mathematical puzzles*, nothing more.

The doctrine that we could not *reason about* the world around us unless we already had the concept of *logic or truth* is nonsense. It is quite the other way around: We could not conceive of *truth* unless we could *pragmatically test hypotheses through action in the world*. *Absolute truth* is a myth, a ghost, a fiction for *logicians and mathematicians*.

As a small step toward testing these grand claims, this article considers the "hot hand" illusion and more generally assumptions about peoples' ability to detect structure in sequences of events. In the following sections we (a) review research on the "hot hand" illusion, (b) explore randomness (or conversely structure or constraint) in relation to the rationality of induction and abduction, (c) present a series of Monte Carlo simulations to illustrate some plausible examples of constraint and evaluate their detectability using various statistical measures, (d) present a series of empirical studies to test people's ability to distinguish whether constraints are present within a sequence of events, and finally (e) we consider the implications for broadening the ecological perspective to consider general properties of cognition.

The hot hand illusion?

The hot hand phenomenon involves two questions about human performance. The first question concerns whether or not the performance of skilled athletes is streaky. The critical question is whether performance on one "trial" (e.g., one shot in a basketball game, one at-bat during a baseball game, or one stroke during a golf match) depends on the successes on prior trials. For example, the claim that a basketball player or baseball player is "hot" indicates that there is an increased probability of success (relative to the player's average) due to a prior string of successes. In this case, the player is experiencing a "hot" streak. Conversely, a player can be in a "cold" streak or "slump" when the probability of success is reduced due to a string of prior failures. The presence of "hot" and/or "cold" streaks suggests that success rates change as a function of sequential dependencies between trials (i.e., events are not independent or the sequence is not generated by a Bernoulli process).

The second question involves people's perception of skilled athletes. Do people believe that performance of athletes is streaky? Is this belief justified, based on the reality of what is actually observed?

Gilovich, Vallone and Tversky's (1985) analysis of basketball shooting brought this issue to the attention of researchers. They analyzed the shooting records of a professional basketball team in the NBA (the Philadelphia 76ers) using serial probabilities and concluded that the likelihood of a successful shot did not necessarily increase following a previous successful shot. They also conducted an analysis on free-throw shooting for the Boston Celtics (another professional basketball team in the NBA), which yielded the same conclusion. Additionally, they reported a controlled shooting experiment with varsity college-level players and concluded that there were no statistically reliable dependencies between the outcomes of successive shots (as measured by conventional statistical tests such as serial correlation and Wald-Wolfowitz Runs Test). Koehler and Conley (2003) evaluated performance from the NBA Long Distance Shootout contest, and Albright (1993) examined hitting performance in professional baseball. Both studies concluded that any streaky behaviors (e.g., strings of successes) were consistent with expectations for a Bernoulli process (i.e., the most valid predictor of success on each attempt is the player's shooting average with no sequential dependencies between attempts).

Despite the conclusion that there is no "streakiness" in basketball shooting, Gilovich et al. (1985) found that many people (including basketball coaches and fans) believed in the "hot hand" phenomenon. Thus, Kahneman (2011) concluded that "the hot hand is entirely in the eye of the beholders, who are consistently too quick to perceive order and causality in randomness. The hot hand is a massive and widespread cognitive illusion" (p. 117). The construct of the "hot hand illusion" suggests that people see patterns (e.g., streakiness) when, in fact, there is no justification in the observed data.

Recently, however, the conclusions of both Kahneman (2011) and Gilovich et al. (1985) have been challenged. For example, do the results for basketball generalize to other sports? In contrast to Gilovich et al., several studies of sports performance have found evidence for streaky performance. For example, Gilden and Wilson (1995) found that performance of novices in golf putting and darts tended to be streaky. Smith (2003; see also Dorsey-Palmateer & Smith, 2004) found evidence for streaky behavior using performance from horseshoe pitchers and tenpin bowlers. In a review and critique of "hot hand" research, Bar-Eli, Avugos, and Raab (2006) found 13 studies that indicated nonstreaky performance (i.e., consistent with a random Bernoulli process) and 11 studies that reported evidence for streaks. However, a major conclusion of this review was that the statistical tests used to assess whether or not performance was streaky were generally weak. For example, Wardrop (1999) and Miyoshi (2000) both found that Gilovich et al.'s statistical tests for detecting streaks were inadequate for the conclusions drawn. Korb and Stillwell (2003) observed the following:

> Gilovich et al. (1985) were incongruously conservative in employing orthodox significance tests to investigate the Hot Hand phenomenon, rather than search for suitable Bayesian inferential procedures. They were surprisingly neglectful in failing to perform any kind of power analysis for the statistical tests they did select. Doing so would have revealed just how weak their tests were, and so also the need for much larger sample sizes than were collected. But worst of all was their claim to have demonstrated the Hot Hand to be a cognitive illusion. In this they succumbed to the Law of Small Numbers themselves, for they were supposing that small, information-poor samples of basketball shooting should support the very same inferences that an arbitrarily large sample could. Nearly a generation of researchers and commentators has subsequently taken it for granted that the Hot Hand phenomenon is nonexistent, and not just in basketball but quite generally. A serious reason to believe that negative conclusion has yet to be offered. (p. 6)

In contrast to beliefs that performance in sporting events is streaky or nonindependent, it has conventionally been assumed that behavior in simple laboratory experiments (e.g., reaction time studies) is independent (i.e., that variations across trials are uncorrelated). However, this belief has recently been challenged (Van Orden, Holden, & Turvey, 2003, 2005) based on an analysis of performance time histories in the frequency domain (i.e., spectral analysis). Independent processes would be expected to yield a frequency pattern consistent with "white noise." However, Van Orden et al. reported an extensive list of human performance studies that find deviations from this expectation. These studies typically find frequency patterns consistent with "pink noise." This is often referred to as "1/f" because the results of the spectral analysis in log-log space show a proportional reduction in the response power with increasing frequency. That is, the magnitude of the signals vary in a way that is inversely proportional to frequency. Van Orden and Holden (2002) concluded that "correlated noise can be observed in all appropriately measured laboratory performances" (p. 101; see also Farrell, Wagenmakers & Ratcliff, 2006; Wagenmakers, Farrell, & Ratcliff, 2005).

The observation that there is structure or dependencies in simple laboratory tasks adds some plausibility to the belief of coaches and sports fans that sports performance might also exhibit dependencies or constraints (i.e., there might be streaks). Or at the least, it makes the alternative hypothesis, that a Bernoulli process (i.e., white noise) is the appropriate model for human performance variance, less plausible.

In addition to the empirical concerns about the generality of Gilovich et al.'s (1985) conclusions that skilled performance is random, there is a fundamental theoretical question about whether the perception of "streakiness" represents an "illusion" or whether it actually represents a "smart mechanism" or a valid or "ecologically rational" conclusion (Brown & Steyvers, 2009; Runeson, 1977; Todd et al., 2012). In "testing" the data, Gilovich, et al. used the logic of null hypothesis testing, where it is assumed that a pattern is random unless proven otherwise. However, we wonder whether this is a reasonable strategy for adapting to a complex world where detection of constraints (structure or invariance) is essential for skillfully realizing affordances. We hypothesize that in such a world, the default hypothesis should be the assumption of structure or regularity unless proven otherwise. We suggest that randomness or independence should be treated as the exception, not the rule. In fact, we would go further to argue that randomness is an especially rare exception in the natural world! Thus, we argue that a "bias" toward the expectation of structure is actually a very smart (or rational) strategy for a system that depends on discovering and leveraging natural constraints in order to survive in a complex world.

Induction, abduction, and randomness

In order to "see the world the way it is" as a prerequisite for successful interaction, it is necessary to discover constraints, patterns, or perhaps laws that structure our experiences. Induction and abduction both refer to the "logic" underlying processes of generalizing from past experience in order to tune to patterns (e.g., invariants) that allow us to anticipate events and that, thus, support skilled interactions with the ecology.

Induction refers to the formal processes for "testing" generalizations or hypotheses about the world. Focus tends to be on identifying the computational limits of inference with respect to absolute truth. A key prescription from inductive logic is that no amount of evidence is sufficient to prove an absolute truth, but a single contradiction is sufficient to show that something is not absolutely true. The implication is that in testing a hypothesis people should focus on disconfirmation rather than confirmation. This raises another concern about human rationality: confirmation bias. A number of studies have shown that people have a "bias" in which they tend to emphasize confirming evidence rather than disconfirming evidence (Wason, 1960). Thus, people violate this prescription for rationality. However, Klayman and Ha (1987) suggested caution in generalizing from the Wason studies. They demonstrated that the Wason task is a special case and showed that a strategy of seeking confirmation can be quite effective under more ecologically representative contexts.

Like induction, abduction is also about making generalizations from past experiences. However, Peirce (1997) framed abduction as a type of common sense or ecological logic, where the purpose was to test beliefs about the world relative to the functional demands of survival. Specifically, Peirce was interested in how our beliefs about the world come to serve the practical demands associated with success or skilled interaction. Peirce was not interested in an absolute truth but rather in a pragmatic truth. This perspective is consistent with Lopes' (1982) suggestion that practically, induction is "a problem in detecting patterns (or non randomness) against a noisy (or random) background" (p. 632). She suggested that signal detection theory is a more useful framework for evaluating the practical value of various decision rules or biases.

Signal detection theory provides an excellent context for differentiating abduction from induction. Essentially, abduction suggests that rather than evaluating rationality relative to the absolute truth of a judgment (true or false), rationality should be evaluated relative to the payoff matrix. The key for success is not to be absolutely correct all the time—which is problematic for an inductive system because there is no absolute "proof" of a hypothesis (e.g., Lopes, 1982, cited Hume, 1748/1977, as showing that induction cannot be justified on logical grounds). Rather, the key for abduction is to set a decision criterion that maximizes the payoffs and/or minimizes the costs over time. This is the pragmatic test of a belief or hypothesis: do choices/actions based on the belief lead to generally satisfying outcomes?

Now consider specifically the problem of discovering patterns in a sequence of events (e.g., determining whether a coin is biased). In this context, any constraints that might exist will be considered signals. Any additional variability will be considered noise. Now consider the benefits of correct detections (i.e., discovering that the coin is biased when it is) relative to the costs of false alarms (i.e., believing a coin is biased when in fact it is unbiased or fair). On one hand, the benefits of recognizing a constraint that exists should be obvious. For example, if the coin is biased so that one side comes up more frequently than another, recognizing this constraint would improve the ability to predict the next flip. On the other hand, what are the costs of a false alarm? If the process is truly random, then any basis for making a guess about the next flip of the coin will typically be as good as any other because the behavior is unpredictable. This suggests that "biasing" the decision criterion to maximize correct detections (at the cost of a high false alarm rate) is a pragmatically smart thing to do. In other words, it is an ecologically rational decision rule because there are benefits to hits but little cost for false alarms. As noted by Scheibehenne, Wilke, and Todd (2011),

> Assuming patterns or regularities in a given environment may be a reasonable default strategy: if there is in fact a pattern, expecting that particular pattern can be advantageous by providing an edge in predicting future events, and if there is no pattern, expecting one will not do worse than any other strategy. (p. 327)

Finally, the construct of randomness itself is quite controversial. In an extensive review of research on humans' abilities to produce and perceive randomness, Nickerson (2002) quoted numerous mathematicians and philosophers to show that even defining what we mean by randomness is controversial. Is it a property of a sequence or a property of the process generating the sequence? Are there any truly random processes in nature or is this simply a useful mathematical abstraction? In summing up the research, Nickerson drew this conclusion:

> I believe that ambiguous or imprecise instruction to participants have been factors in a sufficiently high percentage of experiments on randomness production and perception that the results, in the aggregate, do not constitute a compelling case for the conclusion that people generally have faulty conceptions of randomness. … The results do not rule out the possibility that, in many cases at least, participants' performance was reasonable, according to their interpretation of their task. (p. 353)

Closing the loop with respect to Gilovich et al.'s (1985) research on basketball shooting, the implication of their statistical evaluations (i.e., based on the null hypothesis of a random statistical process—that shooting is an independent process with no more streaks than would be expected for a Bernoulli process) is that the only valid predictor of a shooter's success is his or her shooting average. This means that no other information is relevant. It implies that the ability of the athlete is stationary (not improving or deteriorating over time) and that factors such as the game site, the game situation, the opposing team, and/or the current health of the athlete are all inconsequential.

Alternatively, many expert basketball coaches (e.g., Bobby Knight and Phil Jackson) believe in the hot hand. For example, Phil Jackson once responded to a reporter's question about a deviation from the typical substitution pattern in a game by saying that the player who remained in the game "was hot, and even a fool knows to go with the hot hand" (Medina, 2010). Perhaps we should trust people with a history of success. Perhaps the null hypothesis should be that shooting behavior is constrained. It may be constrained by many conflicting factors so that in the long run the sequences of successes and failures may be consistent with many of the statistical properties of a Bernoulli process. However, in the short run a string of successes or failures may be indicative of local constraints (e.g., a weak defender or fatigue) such that shifting decision criterion (e.g., feeding the hot hand) may be an ecological valid basis for making a coaching decision.

Monte Carlo simulations

At this point, we do not know whether or not actual performance of athletes is constrained (i.e., statistically random) as there is controversy over the logic of the statistical tests used to make this judgment. In this section, we generate sequences with known constraints that are somewhat plausible with respect to human performance and then we "test" the sequences using several common statistical methods to gauge the sensitivity of the statistical methods. Thus, the question asked in this section is, if there are constraints on the sequence, will these constraints be detected given the commonly used statistical methods?

Simulation method

Binary shooting sequences were generated using Monte Carlo simulations in order to represent plausible real-world constraints that could impact a player's shooting performance. The criteria that guided the different types of constraints were based on the dependencies of trials

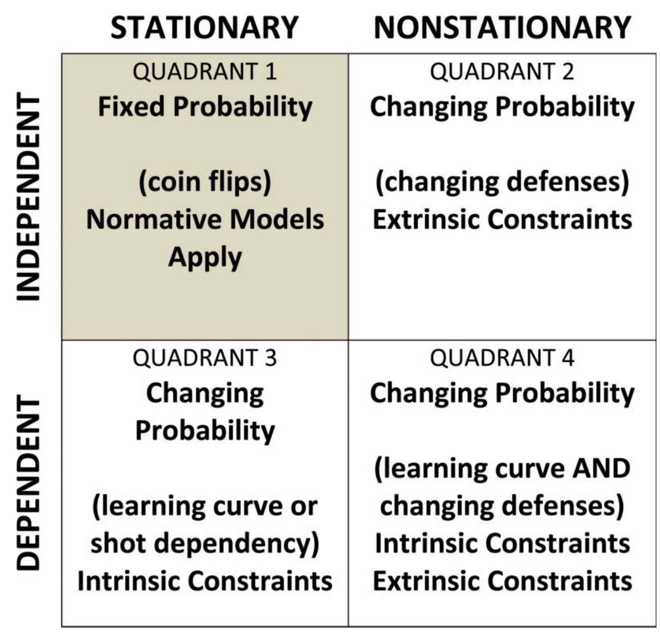

Figure 1. This diagram illustrates four types of processes as a function of whether successive trials are independent and if the generating rules are stationary. The constraints associated with Quadrants 2, 3, and 4 were chosen to reflect plausible constraints on sports performance.

coupled with the stationary properties of the rules governing the sequence, as illustrated in Figure 1. The vertical dimension in Figure 1 reflects whether performance on one trial depended in any way on previous performance (e.g., sequential dependencies, learning, etc.). The horizontal dimension in Figure 1 reflects whether the rules, parameters, or algorithms for determining performance on a trial were fixed (stationary) or variable (nonstationary).

Quadrant 1

The binary sequences generated from Quadrant 1 represent independent, stationary processes. These Bernoulli sequences correspond to an unconstrained process (i.e., probabilistic process) where the fixed probabilities of success for each of the three simulations were .30, .50, and .80.

Quadrant 2

The simulations in Quadrant 2 represent sequences composed from two different Bernoulli processes (i.e., an independent, nonstationary process). One process has a probability of success of .60 (hot streak) and the other has a probability of success of .20 (cold streak). This was representative of a scenario where a player had different success rates in different contexts (e.g., as a function of the quality of defense or the game location—home vs. away). Three simulations were generated with the probability of switching equal to .10, .25, and .50. For example, if the player was currently hot (.60), then .10 would be the probability that the player would be cold (.20) on the next trial. Thus, the process with the lowest probability of switching (.10) would have relatively longer hot or cold streaks.

Quadrant 3

The simulations generated in Quadrant 3 represent different variations of dependent, stationary processes. Three different constraint types were generated with either local or global shot dependencies. The first constraint type had trials constrained by local shot dependencies where the success of a given trial was dependent upon the success of the immediately prior shots. One set of sequences was generated using the rule that if the preceding trial was a success the probability of making the next trial became .70 and if it were a miss the probability became .30. The second set of sequences had the success of any given trial based on the success of the preceding five trials: if two or fewer successes, p(success) = .30; if three successes, p(success) = .50; if four or more successes, p(success) = .70. These would represent the classical model of streakiness in which future success was a function of success on recent previous trials.

The second constraint type had trials constrained by global shot dependencies where the success of a given trial was dependent upon the number of prior shots generated. This was meant to represent a player whose performance is based on a learning curve where the probability of success increases as the number of shots taken increases (i.e., "practice makes perfect"). In this practice-based global shot dependency condition there were three different sequence types generated where the probability of success increased exponentially in relation to the number of prior shots taken. All sequences in this practice-based global shot dependency condition had an initial probability of success equal to .00 and asymptotic probability of success equal to .80. The three sequence types varied only in the learning rate (k), which was set to .001, .003, or .005.

The third constraint type had trial success dependent on the success of all prior shots taken. This was meant to represent a player whose performance is based on a learning curve where the probability of success increases as the number of *successful* shots taken increases (i.e., "perfect practice makes perfect"). In this success-based global shot dependency condition there were six different sequence types generated where the probability of success increased exponentially in relation to the number of successful shots taken.

Three of the success-dependent sequence types had the initial probability of success equal to .33 and the asymptotic probability of success equal to .61, whereas the other three simulations had the initial p(success) = .20 and the asymptotic p(success) = .80. The initial p(success) of .33 was determined by looking at the lowest percentage of field goals made by an NBA player in a regular NBA season. The same method was used for the asymptotic probability of success except that the player with the highest percentage of field goals made was used. A total of seven regular NBA seasons were selected and the percentage of field goals made by the worst player from each season and best player from each season was found. After averaging the seven worst players together and seven best players together, the worst players averaged p(success) of .33, whereas the best players averaged a p(success) of .61.

In order for a player to be used in the computation the player must have taken at least 246 shots in that season (i.e., on average 3 shots per game in an 82-game season), thus excluding players with nonrepresentative shooting performance. There were three learning rates used within these two ranges of initial and asymptotic probabilities of success, k = .001, .003, or .005, thus resulting in six different sequence types being generated.

Quadrant 4

In Quadrant 4 there was only one simulation generated. This process combined the practice-based exponential learning rate used in Quadrant 3 (initial success probability = .20; asymptotic success probability = .60; learning rate = .002) and the alternating probability of success used in Quadrant 2. Therefore, the actual probability was increased by .10 above or below the probability computed from the learning function. As with Quadrant 2 the alteration between +.10 or −.10 was based on a fixed probability equal to .10. This resulted in persistent hot (+.10) or cold (−.10) streaks on top of a typical learning curve.

Simulation analysis

To determine whether the constraints implemented in these simulations could be detected by common statistical methods, the Wald-Wolfowitz Runs Test and the slope of the frequency response function (Beta) resulting from Fourier Analysis were used. In order to determine Beta, the binary sequences were normalized and converted to the frequency domain using a Fourier analysis package in MATLAB. A regression line was then fit to the spectral data in log-log space. The slope of this regression line provided the beta value for that sequence.

Ten binary sequences were generated for each of the simulation types, each with 1,024 data points (i.e., shots/trials). This yielded 10 beta values for each of the 18 simulation types. The same binary strings used to find these beta values were also used in the Wald-Wolfowitz Runs Test analysis using Microsoft Excel. The z statistic was used as the conventional test of randomness for the Runs Test, and two-tailed t tests were used to evaluate whether the slopes in these simulations were reliably different from zero. Thus, this allowed testing for deviations from the expectations of a Bernoulli process for the Runs Test (large absolute z score) and/or spectral analysis (slopes that are significantly different than zero). Furthermore, all of the simulation types were analyzed using 512, 800, or 1,024 data points from the originally generated 1,024 data point sequences.

Simulation results

Table 1 summarizes the results for the 18 simulations that were conducted across the four different quadrants. The table includes the overall mean performance (percent success), the results of the Wald-Wolfowitz Runs Tests using various numbers of data points, and the mean slopes (betas) and two-tailed t test results from the spectral analysis using various numbers of data points.

Quadrant 1

The p(hit) = .80 condition was the only simulation in this quadrant to show deviation from the expectations of a Bernoulli process, which only occurred for the spectral analysis using the weakest sample size of 512 data points. Note that this is a false rejection because all the sequences in this quadrant were generated using a Bernoulli process. Thus, for all the other cases in this quadrant, the null results for both the Runs Test and the Slope Test were accurate in retaining the null hypothesis that these sequences did not differ from the expectations for a Bernoulli process.

Quadrant 2

In this quadrant, the simulation with a 50% chance of alternating between hot and cold streaks resulted in null results for both tests and all sample sizes, indicating that these sequences did not differ from the expectations for a Bernoulli process. The other two constraint conditions did result in significant deviations from the expectations for a Bernoulli process in all cases except for the 25% chance of switching condition when using the Runs Test with the smallest sample size. Again, this is a miss because these sequences were not

Table 1. The overall mean performance (% success), the results of the Wald-Wolfowitz Runs Tests analyzing various amounts of data points, and the mean slopes (betas) and two-tailed *t* test results from the spectral analysis using various data points.

			Runs test			Frequency analysis beta slope					
	N = 1,024		*N* = 512	*N* = 800	*N* = 1,024	*N* = 512			*N* = 1,024		
	M	*SD*	$\lvert Z \rvert$	$\lvert Z \rvert$	$\lvert Z \rvert$	*M*	*SD*	*t*	*M*	*SD*	*T*
Quadrant 1											
Bernoulli processes											
p(hit) = .30	0.306	0.013	0.04	0.18	0.44	0.02	0.06	0.87	0.02	0.04	1.50
p(hit) = .50	0.500	0.018	0.26	0.06	0.02	0.03	0.05	1.94	0.01	0.06	0.26
p(hit) = .80	0.806	0.009	0.62	0.57	0.05	−0.05	0.06	−2.58*	−0.01	0.07	−0.63
Quadrant 2											
p(hit) = .20 & p(hit) = .60											
10% chance of alternation	0.419	0.018	3.04*	3.87*	4.18*	−0.23	0.07	−10.75*	−0.25	0.07	−11.93*
25% chance of alternation	0.408	0.014	1.81	2.18*	3.08*	−0.15	0.09	−5.15*	−0.18	0.07	−8.54*
50% chance of alternation	0.401	0.014	0.01	0.21	0.02	0.00	0.06	−0.11	0.01	0.05	0.62
Quadrant 3											
Shot dependencies											
Last 1-shot dependency	0.498	0.020	8.89*	11.10*	12.67*	−0.55	0.12	−14.73*	−0.56	0.08	−23.54*
Last 5-shot dependency	0.422	0.034	3.66*	4.40*	5.11*	−0.26	0.05	−15.46*	−0.26	0.07	−11.39*
Practice-dependent learning											
k = 0.001	0.292	0.011	0.84	2.54*	3.79*	−0.05	0.07	−2.30*	−0.08	0.04	−6.94*
k = 0.003	0.556	0.015	3.10*	4.68*	5.70*	−0.14	0.08	−5.71*	−0.10	0.05	−6.72*
k = 0.005	0.642	0.008	4.04*	5.33*	5.74*	−0.14	0.05	−9.33*	−0.08	0.04	−6.41*
Performance-dependent learning											
p(hit) = .33 & p(hit) = .61											
k = 0.001	0.380	0.019	0.19	0.16	0.40	−0.01	0.08	−0.35	−0.03	0.06	−1.37
k = 0.003	0.459	0.022	0.33	0.64	0.66	−0.03	0.07	−1.15	−0.03	0.06	−1.60
k = 0.005	0.499	0.020	0.33	0.48	0.75	−0.05	0.06	−2.37*	−0.02	0.05	−1.64
p(hit) = .20 & p(hit) = .80											
k = 0.001	0.273	0.013	0.17	0.16	0.35	−0.01	0.08	−0.40	−0.02	0.07	−0.80
k = 0.003	0.425	0.016	0.41	1.17	2.10*	−0.03	0.08	−1.35	−0.03	0.06	−1.86
k = 0.005	0.529	0.022	1.31	3.11*	4.09*	−0.11	0.06	−5.63*	−0.09	0.05	−5.53*
Quadrant 4											
Simple learning curve and p(hit) +/−10%											
k = 0.002 +/−10%	0.391	0.022	1.98 *	2.64 *	3.18*	−0.16	0.05	−10.23*	−0.13	0.05	−8.28*

Note. N refers to the number of data points in a given sequence. *N* = 512 and *N* = 800 indicate that the beginning 512 and 800 (respectively) data points from the 1,024 data-point sequence were used.
*$p < .05$ for one-tailed *z* test for runs. *$p < .05$ for two-tailed *t* test for slope (beta).

generated by a stationary, Bernoulli process. The overall result, however, is that the switching between two Bernoulli processes to produce hot and cold streaks was generally detected as a deviation from a strictly Bernoulli process by both the Runs Test and the Slope Test (as long as the sample size was adequate and the switching rate was not 50%).

Quadrant 3

Three different types of constraints were used for the simulations in Quadrant 3: local dependency, trial-dependent learning, and success-dependent learning. For the *local-dependency conditions* and the *trial-dependent learning conditions*, all tests except one correctly detected the deviations from a strictly Bernoulli process (i.e., detected the constraints). Not surprisingly, the one exception was with the smallest sample size.

The results for the *success-dependent learning conditions* were surprising. In the three conditions that were designed so that the range of improvement reflected lower and upper boundaries on success rates that were representative of NBA shooting performance, both tests at all sample sizes failed to detect the presence of a constraint (i.e., the results were not significant), with one exception. The frequency analysis found a significant slope effect for the highest learning rate using the smallest sample size. For the sequences with larger performance ranges (i.e., p(success) = .20 to .80) the results were mixed. The constraints were almost exclusively missed (i.e., null results) for the two lowest learning rates except for one test using the largest sample size. However, with the highest learning rate, both tests correctly detected the deviation from a strictly Bernoulli process (i.e., detected the constraint) with the exception of one test using the smallest sample size.

Quadrant 4

In this quadrant the presence of the constraints was detected by both tests with all sample sizes.

Discussion of simulation results

Overall, both statistical tests were generally accurate in discriminating a purely Bernoulli process from a constrained process, particularly when the sample size was large. That is, the tests yielded null results in Quadrant 1, when the simulations were based on a Bernoulli process, and they yielded significant results for most of the simulations in the other quadrants, which were not stationary Bernoulli processes, as long as the sample sizes were large (i.e., $\geq$ 800).

The interesting exception was the success-based learning conditions with the narrow performance range in Quadrant 3. This is particularly interesting because the narrow range of performance was determined based on the observations of actual shooting ranges for NBA players. Thus, in this condition, although the sequences were constrained, the constraints were not detected by either statistical method with the sample sizes used. So, here is at least one example where there are known constraints on the sequences, but these constraints are not detected using the logic of null hypotheses testing. Although the hypothesis of a Bernoulli process cannot be rejected based on these tests, the null results clearly do not prove the absence of constraint!

Empirical studies of human judgment

In this section, five experiments are presented in which participants were asked to discriminate between two different processes for generating strings of shooting performance. In each experiment, participants had access to a sequence of 800 shots that were generated by one of two different processes. The nature of the different constraints was described to the participants in colloquial terms so that they had some sense of what type of constraints might be shaping performance for each of the two processes. Note that this contrasts to many previous studies where participants were typically asked simply whether or not the process was random. As both Nickerson (2002) and Lopes (1982) noted, this question is ill posed in that any specific sequence is as likely as any other specific sequence if the process is random. As noted earlier, Lopes suggested that a more realistic framework for assessing judgments about strings is the signal detection paradigm in which participants are asked to discriminate between two different specific processes. In this case, the signals from a random process, in which any sequence is possible, would result in a distribution that would completely overlap with any alternative process. Although it doesn't make sense to ask whether any particular sequence is random, it does makes sense to ask whether a particular sequence or class of sequences (e.g., a string of 100 straight hits without a miss) is more likely from a random generator or another specific process.

General method

Participants

Each of the five empirical experiments had 16 people participate. Participants were recruited from the undergraduate psychology students at Wright State University who participated in this experiment in partial fulfillment of their obligations to obtain required research credit.

Equipment

All experiments were carried out using an Apple iMac and the program was developed through the Java runtime environment. A standard optical mouse was used to control the display and select the desired options. The interface, illustrated in Figure 2, included a window through which the participant could view one binary sequence at a time. Each sequence had a total of 800 events (i.e., successes and misses). Participants had a sliding window that allowed them to only view 16, 4, or 1 event of the sequence at a time.

The binary events were either an empty circle to represent a missed shot or a basketball icon to represent a successful shot. A triangular slide control at the top of the screen allowed the participants to scan the entire sequence of 800 events by sliding the viewable event window of 16, 4, or 1 throughout the sequence. A number was displayed above the slider to indicate the trial

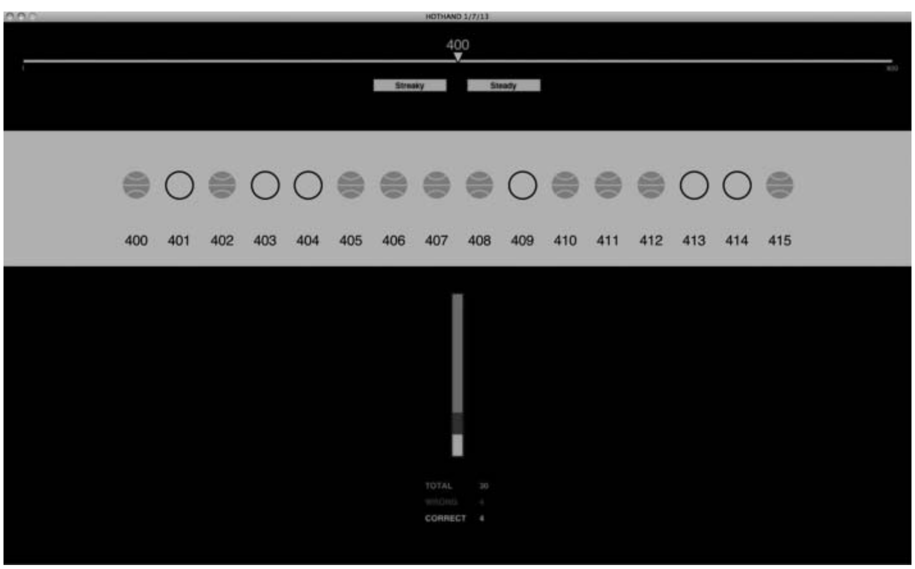

Figure 2. The interface with Window Size = 16. The slider at the top allowed participants to scan a sequence of 800 shots, 16 at a time.

number for the first visible event in the window. After examining the sequence participants entered a response using one of two control buttons above the display window. A display below the viewing window provided both graphical (bar graph) and digital feedback about the number of correct and incorrect answers. This updated immediately following each response.

Procedure

All five empirical experiments followed the same procedure. For training, participants were allowed six practice trials where they were given one of each type of sequence for each of the three viewing window sizes (i.e., 16, 4, and 1). Participants were able to ask for any clarifications needed during that time. Every effort was given to ensure that the participants understood the nature of the task they were performing. Participants then had the opportunity to repeat the practice trials until they understood the nature of the task. After the practice trials participants were told that no more assistance would be given and that they would not be able to ask further questions once the experimental trials started.

All experiments involved three blocks of 30 trials each. Window size was manipulated in a fixed order across blocks. Window Size = 16 was used for the first block, Window Size = 4 was used for the second block, and Window Size = 1 was used for the third block. The 30 trials in each block included 15 presentations of each type of sequence being discerned. Trials were presented in a random order that did not change between participants.

At the beginning of every training and experimental trial the viewing window was positioned in the middle of the 800-event sequence. Participants were allowed to freely explore the sequences by scrolling left and right until a choice was made. Participants entered their decision by clicking the appropriate response button, which resulted in immediate feedback and made the next sequence available. The completion time for trials was not constrained (participants could take as long as needed).

Dependent variables

The dependent measures included the proportion of correct responses (categorized in a signal detection matrix as proposed by Green & Swets, 1966/1974), an estimated *d'* for each participant, a response bias value for each participant, the total time taken per trial, and the amount of scanning motion (*MSE* or variance of the scanning motion). A hierarchical Bayesian modeling procedure, as described by Rouder and Lu (2005), was used to estimate *d'* values and response bias or *c* values for each participant as a function of the numbers of hits and false alarms in each block of 30 trials as well as the overall group-level parameters.

Experiment-specific methods

Experiment 1

The participants were asked to distinguish between "poor" and "good" shooters. The binary sequences used were all generated from a Bernoulli process where half of the sequences represented poor shooters with a shooting percentage of p(success) = .20 and the other half represented good shooters with a p(success) = .60 (based on Artis

Gilmore's actual NBA career field goal percentage of 59.9). The response buttons for this experiment were thus labeled "Poor" and "Good" to indicate the different types of sequences. The shooting sequences used were generated identical to the Bernoulli simulations in Quadrant 1 and represent independent, stationary processes.

Experiment 2

Participants were asked to distinguish between "streaky" and "steady" shooters (the response buttons were labeled as such). The binary sequences generated for the streaky shooters used the constraints discussed in Quadrant 2, where any given shot generated had a p(success) = .20 or .60. There was a 10% chance of alternating between these two potential shooting probabilities independent of the success of any given shot. Thus, the streaky shooter represented an independent, nonstationary process. The shooting sequences used for the steady shooters were generated identical to the Bernoulli simulations in Quadrant 1 and represent independent, stationary processes with a p(success) = .50.

Experiment 3

Participants were asked to discriminate between "streaky" and "steady" shooters. The binary sequences generated for the streaky shooters in this experiment used the local-shot-dependency constraint, where the success of any given shot was based on the success of the preceding five trials: if two or fewer successes, p(success) = .30; if three successes, p(success) = .50; if four or more successes, p(success) = .70. This is consistent with Quadrant 3 in Figure 1, representing a dependent, stationary process. The shooting sequences used for the steady shooters were generated by independent, stationary processes with a constant p(success) = .50.

Experiment 4

Participants were asked to discriminate between "improving" and "steady" shooters. The binary sequences generated for the improving shooters in this experiment used the trial-dependency constraint, where the probability of success increases as the *number of shots taken* increases. The success of any given shot was based on a learning curve rate of .003, starting with a p(success) = 0.1 and an asymptote at p(success) = 0.6. This was similar but not identical to Quadrant 3 in Figure 1 representing a dependent, stationary process. The shooting sequences used for the steady shooters were generated by independent, stationary processes with a constant p(success) = .50.

Experiment 5

Participants were asked to discriminate between "improving" and "steady" shooters. The binary sequences generated for the improving shooters in this experiment used the success-dependent learning constraint, with a learning rate = .005, a starting p(success) = .33, and an asymptote at p (success) = .61 (recall these parameters were based on actual NBA shooting performance). This is consistent with Quadrant 3 in Figure 1, representing a dependent, stationary process. The shooting sequences used for the steady shooters were generated by independent, stationary processes with a constant p(success) = .44. This success rate was computed by taking seven regular NBA seasons and averaging the field goal percentages of the median player from each season.

Results

Experiment 1

The results for mean hit rates, false alarm rates, and adjusted *d'* as a function of Window Size are presented in Table 2. The Bernoulli sequences of good and poor shooters were by far the easiest discriminations for participants to make. In fact, 6 of the 15 participants correctly identified the good shooter from the poor shooter for all trials in all blocks without making any errors. If participants were at chance performance, their *d'* would be 0. To test whether $d' = 0$ was plausible, we examined the posterior distributions from the hierarchical Bayesian model of adjusted *d'*, which indicated essentially no probability that $d' \leq 0$. Pairwise comparisons of the *d'* posterior distributions as a function of Window Size show no reliable differences.

Experiments 2–4

Experiments 2–5 were most critical for the question of whether participants could discriminate Bernoulli random sequences (generated by a Bernoulli process with p(success) = .5)

Table 2. Mean percentage of hit rates, false alarm rates, and adjusted *d'* as a function of window size. All of the posterior means are from the Bayesian estimates.

		Hit rate		False alarm rate		Adjusted *d'*	
		M	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1							
Bernoulli processes good/poor							
	Window = 16	0.96	0.01	0.04	0.02	6.00	1.00
	Window = 4	0.96	0.01	0.04	0.01	5.50	0.00
	Window = 1	0.93	0.04	0.04	0.01	4.80	0.80
Experiment 2							
10% alteration p(hit) = .2 & .6							
	Window = 16	0.72	0.19	0.39	0.20	1.00	0.20
	Window = 4	0.60	0.24	0.37	0.13	0.60	0.20
	Window = 1	0.67	0.21	0.37	0.17	0.90	0.20
Experiment 3							
5 shot dependency							
	Window = 16	0.78	0.20	0.28	0.16	1.60	0.30
	Window = 4	0.70	0.18	0.29	0.16	1.20	0.20
	Window = 1	0.63	0.17	0.25	0.15	1.10	0.20
Experiment 4							
Simple learning curve							
	Window = 16	0.86	0.08	0.13	0.12	2.40	0.20
	Window = 4	0.84	0.12	0.21	0.19	2.20	0.30
	Window = 1	0.69	0.18	0.24	0.17	1.30	0.20
Experiment 5a							
Performance-dependent learning curve							
	Window = 16	0.64	0.10	0.33	0.20	0.82	0.17
	Window = 4	0.67	0.12	0.31	0.17	0.96	0.16
	Window = 1	0.61	0.15	0.34	0.13	0.70	0.13
Experiment 5 Recoded							
Performance-dependent learning curve							
	Window = 16	0.65	0.10	0.32	0.21	0.86	0.17
	Window = 4	0.72	0.12	0.29	0.16	1.16	0.14
	Window = 1	0.61	0.15	0.34	0.13	0.70	0.12

Note. Mean (*M*) and Standard Deviation (*SD*) were taken across participants.

from sequences with constraints. Experiments 2–4 had participants viewing constraints that were reliably detected in the Monte Carlo simulations by the Runs Test and spectral analysis. All three experiments showed the posterior distributions of adjusted *d'* across all window sizes to have effectively zero probability that $d' \leq 0$, indicating that humans were able to discriminate between the Bernoulli process and each of these three different constraint types.

Experiment 5

The most interesting results came from the constraint used in Experiment 5. Recall that the Runs Test and spectral analysis did *not* detect the success-dependent learning constraint using parameters based on actual NBA performance as being significantly different from a Bernoulli process. The analysis of human performance, however, showed the distributions of adjusted *d'* across all window sizes to have no posterior probability of $d' \leq 0$, indicating that humans were able to make the discrimination despite the fact that the statistical tests were not significant (see Table 2).

Experiment 5 recoded

Recoding method

Sequences generated by a Monte Carlo process will sometimes deviate from the expectations of a typical sequence (based on the parameters generating it). Thus, although the Monte Carlo simulation was tuned to generate streaky sequences, some of the sequences generated could be more consistent with the steady shooter (i.e., a nonstreaky Bernoulli sequence) and vice versa. Bayesian model comparisons were conducted for all of the empirical experiment trials ($N = 90$) used in Experiment 5. This was done to verify that the data were more representative of the model that generated it (e.g., streaky) than the alternative (e.g., white noise).

Both models that were used to examine the data represented optimal performance assuming full knowledge of the task parameters. One model assumed a Bernoulli process where the p(hit) = .44 (a "steady" shooter), and the other model assumed the success-dependent learning process (a "streaky" shooter), where the moving probability of success was assumed to be computed using the following formula, where I = initial p(hit), A = asymptotic p(hit), k = learning rate, S = sum of all prior hits:

$$p(\text{hit}) = I + \left((A - I) \times \left(1 - e^{(-(k) \times S)}\right)\right).$$

For sequences (considered a trial in the experiment) with a Bayes factor of less than one, the trial was recoded and the correct answer was switched to the alternative option, so a streaky generated sequence would get recoded for analysis as being a steady sequence and vice versa. Only three sequences/experimental trials had a Bayes factor of less than one. One trial was recoded from a Bernoulli-generated sequence (steady) to a success-dependent learning sequence (streaky) in Window Size 16, and two trials were recoded from success-dependent learning sequences to Bernoulli-generated sequences, one in Window Size 16 and one in Window Size 4.

Recoded results

The posterior probability that the overall *d'* across all window sizes after recoding is larger than before recoding is 0.64 and the correlation between participant *d'* before and after recoding is 0.95, which is not strong evidence for a difference in *d'* after the three experimental trials were recoded. The posterior distributions of adjusted *d'* across all window sizes showed no posterior probability that $d' \leq 0$ after the three experimental trials were recoded (see Table 2), which still indicates that humans were able to make the discrimination that the statistical tests could not.

Summary and conclusions

We believe that this work has theoretical, methodological, and empirical implications relative to understanding human rationality. The theoretical point is that, as with perception, it is important to ground our theories of human rationality in the functional dynamics of tuning to and adapting to the affordances in the ecology (e.g., see Flach & Voorhorst, 2016; Todd et al., 2012). This suggests that a pragmatic approach consistent with Peirce's (1997) logic of abduction may be a useful context for gauging the ecological rationality of human choices and judgments (as opposed to context-independent logics, such as formal models of deduction and induction). With specific reference to the detection of constraints associated with sequences of activity, we suggest that the assumption of structure, unless proven otherwise, is a smart or ecologically valid attitude for people (e.g., coaches) to adopt.

The methodological issues associated with gauging human intuitions about randomness have been well summarized by Lopes (1982) and Nickerson (2002). Randomness is an abstract, ill-defined mathematical construct that must be approached cautiously when describing sequences. Asking people to judge randomness will be particularly sensitive to demand characteristics. Thus, we recommend using tasks that have some functional significance to participants. For example, Nickerson noted that when participants are put in a competitive situation, where generating unpredictable sequences provides an advantage, the sequences that they generate tend to meet the mathematical criteria for randomness.

Additionally, in scoring or gauging human performance it may be important to assess performance against pragmatic constraints (e.g., the payoff matrix associated with signal detection) rather than with respect to an experimenter's abstract, context-free criterion for "truth" (e.g., conformity with some normative mathematical model). As with research in perception–action, we suggest that it is essential to use experimental contexts that are representative of the functional problems people need to solve in natural environments, with representative constraints on the available information and the consequences (e.g., avoiding collision rather than passively judging absolute physical values such as speed and distance). In making coaching decisions about when to substitute for a player, we suggest that in addition to the shooting quality (e.g., shooting percentage), situational (e.g., opposition strengths and weaknesses), emotional (e.g., player's motivation or confidence), and social constraints (e.g., team cohesion) may be important considerations. Although expert coaches may not be able to explicitly specify all the constraints that are influencing their choices, belief in the hot hand may reflect their sense that players' performance is not simply a function of chance.

Finally, the empirical results suggest that, given information about the kind of patterns they are trying to detect, people seem to be reasonably good at discriminating between constrained and random sequences. For the particular discriminations that we tested, the human performance was generally consistent with the standard statistical tests and in at least one case, the human performance seemed to be better than the standard null hypothesis testing typically used to judge randomness. Note that the performance was not "optimal" relative to a Bayesian model with complete knowledge of the alternative generation algorithms. However, this is not surprising given the limited practice provided to our participants.

Acknowledgments

Thanks to Arisara Jiamsanguanwong, who helped to explore and develop the models for the Monte Carlo simulations while a visiting researcher at Wright State University supported through the Tokyo Institute of Technology. Also, thanks to Lucas Lemasters, who helped to run some of the Monte Carlo models. Randy Green wrote the software for the experimental evaluations. Thanks also to John "Jay" Holden, who helped with the Fourier Analyses. Parts of this research were reported in Jeh Cooper's doctoral dissertation and Taleri Hammack's master's thesis, which were submitted to the graduate school at Wright State to fulfill degree requirements.

References

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

Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of "hot hand" research: Review and critique. *Psychology of Sport and Exercise*, *7*(6), 525–553.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*, 49–67.

Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *American Statistician*, *58*, 38–45.

Farrell, S., Wagenmakers, E. J., & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean? *Psychonomic Bulletin & Review*, *13*(4), 737–741.

Flach, J. M., & Voorhorst, F. A. (2016). *What matters?* Dayton, OH: Wright State University Library. Retrieved from http://corescholar.libraries.wright.edu/books/127/

Gibson, J. J. (1982). The useful dimensions of sensitivity. *American Psychologist*, *18*, 1–15. Reprinted in E. Reed & R. Jones (Eds.), *Reasons for realism* (pp. 350–373). Hillsdale, NJ: Erlbaum. (Original work published 1963)

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum. (Original work published 1979)

Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. *Psychonomic Bulletin & Review*, *2*, 260–265.

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

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Robert E. Krieger. (Original work published 1966)

Hume, D. (1977). *An enquiry concerning human understanding*. Indianapolis, IN: Hackett. (Original work published 1748)

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus and Giroux.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Koehler, J. J., & Conley, C. (2003). The 'Hot Hand' myth in professional basketball. *Journal of Sport & Exercise Psychology*, *25*, 253–259.

Korb, K. B., & Stillwell, M. (2003, July). *The story of the Hot Hand: Powerful myth or powerless critique?* Presented at the International Conference on Cognitive Science, Sydney, Australia.

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.

Medina, M. (2010, October 28). Round-the-clock purple and gold [Web log post]. Retrieved from http://lakersblog.latimes.com/lakersblog/2010/10/phil-jackson-acknowledges-challenge-in-fitting-in-sasha-vujacic-into-rotation.html

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

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330–357.
Peirce, C. S. (1997). *Pragmatism as a principle and method of right thinking: The 1903 Harvard lectures on pragmatism*. New York: State University of New York Press. Retrieved from http://philpapers.org/rec/PEIPAA
Rouder, J., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604.
Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, *18*, 172–179.
Scheibehenne, B., Wilke, A., & Todd, P. (2011). Expectations of clumpy resources influence predictions of. *Evolution and Human Behavior*, *32*, 326–333.
Smith, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin & Review*, *10*, 753–758.
Todd, P. M., Gigerenzer, G., & The ABC Research Group. (2012). *Ecological rationality intelligence in the real world*. New York, NY: Oxford University Press.
Van Orden, G. C., & Holden, J. G. (2002). Intentional contents and self control. *Ecological Psychology*, *14*, 87–109.
Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self- organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*, 331–350.
Van Orden, G. C., Holden, J. G., & Turvey, M. (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology: General*, *134*, 117–123.
Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, *134*, 108–116.
Wardrop, R. L. (1999). *Statistical tests for the Hot-Hand in basketball in a controlled setting*. Retrieved from http://pages.stat.wisc.edu/~wardrop/papers/tr1007.pdf
Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.