2015

Detecting Structure in Activity Sequences: Exploring the Hot Hand Phenomenon

Taleri Lynn Hammack

Wright State University

DETECTING STRUCTURE IN ACTIVITY SEQUENCES: EXPLORING THE HOT HAND PHENOMENON

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

TALERI HAMMACK
B.S., University of Idaho, 2012

2015
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

August 20, 2015

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Taleri Hammack ENTITLED Detecting Structure in Activity Sequences: Exploring the Hot Hand Phenomenon BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Masters of Science.

John Flach, Ph.D.
Thesis Director

Scott Watamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

Committee on
Final Examination

John Flach, Ph.D.

Kevin Bennett, Ph.D.

Joseph Houpt, Ph.D.

Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Hammack, Taleri Lynn. M.S. Human Factors and Industrial/Organizational Psychology, Department of Psychology, Wright State University, 2015. Detecting Structure in Activity Sequences: Exploring the Hot Hand Phenomenon.

Can humans discriminate whether strings of events (e.g., shooting success in basketball) were generated by a random or constrained process (e.g., hot and cold streaks)? Conventional wisdom suggests that humans are not good at this discrimination. Following from Cooper, Hammack, Lemasters, and Flach (2014), a series of Monte Carlo simulations and an empirical experiment examined the abilities of both humans and statistical tests (Wald-Wolfowitz Runs Test and 1/f) to detect specific constraints that are representative of plausible factors that might influence the performance of athletes (e.g., learning, non-stationary task constraints). Using a performance/success dependent learning constraint that was calibrated to reflect shooting percentages representative of shooting in NBA games, we found that the conventional null hypothesis tests were unable to detect this constraint as being significantly different from random. Interestingly however, the analysis of human performance showed that people were able to make this discrimination reliably better than chance. Hence, people may also be able to detect patterned/constrained processes in a real-world setting (e.g., streaks in basketball performance), thus supporting the belief in the hot hand.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to acknowledge and thank all of my committee members, Dr. Kevin Bennett, Dr. John Flach, and Dr. Joseph Houpt. Thank you all for your time, input, and guidance. I would especially like to thank my advisor Dr. John Flach for his insights and commitment to this work. His guidance on this project, and more broadly my graduate school career, has been invaluable. His passion for continual exploration of cognitive systems is contagious and I look forward to earning my Ph.D with him.

I would also like to thank my parents and especially my grandmother. Their encouragement and support of me and my education are fundamental to the person I have become and the work I have accomplished. Last, but certainly not least I would like to thank my wife. Her unwavering dedication to me and all of my academic quests reminds me everyday why I chose to spend the rest of my life with her. She continues to make me a better, wiser, and stronger person.

1. INTRODUCTION

1.1. Heuristic: Bias or Smart Mechanism

Should the use of heuristics be viewed as biased, suboptimal judgment strategies, or alternatively viewed as smart, adaptive decision making mechanisms? A considerable portion of the decision making literature is concerned with finding ways to mitigate the biasing effects of heuristics that too often lead us astray from the normative solution. However, Shah and Oppenhiemer (2008) proposed that heuristics should be viewed as a small set of effort-reducing principles rather than broad cases of decision making strategies that often times lead to cognitive failures.

Gigerenzer and Gaissmaier (2011) point out that because heuristics are mental short-cuts where people ignore certain information in order to reduce cognitive effort, it would seem like the probability for erroneous decisions is much higher since potentially critical information has been overlooked. However, they argue that judgments can actually be more accurate when some information is overlooked, as opposed to strategies based on exploring all of the information. Lopes and Oden (1991) further argue that when the use of heuristics tends to appear erroneous is when our general knowledge of how the world works is exploited in a narrow laboratory setting.   Take for example the infamous Linda problem (Tversky & Kahneman, 1974):

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of

the following two alternatives is more probable? (1) Linda is a bank teller or (2) Linda is a bank teller and active in the feminist movement.

A majority of subjects who have been given this problem say that it is more likely that Linda is a bank teller *and* active in the feminist movement (they presumably rely on the representativeness heuristic in assessing the probabilities of each option). This answer is typically viewed as incorrect, because mathematically the p(A) is not greater than the p(A&B).

Now take this same problem and apply it to a real-world scenario. If given this same description about an individual who was randomly sampled from a group of people, the only supporting information for Linda being a bank teller is that she is a female (which most bank tellers are). There is nothing else in this description that would activate features associated with a bank teller. However, there are multiple statements that would suggest Linda's potential involvement with a feminist movement (e.g., deeply concerned with discrimination and social justice, outspoken, etc.). An ecologically valid answer, since Linda would be a bank teller in either option, would be that Linda is also active in the feminist movement, based on the amount of supporting evidence in the description.

Lopes and Oden (1991) argued that:

> *…the normative criteria held in such high esteem by psychologists are often irrelevant for human action…the apparent failures people manifest in many laboratory tasks actually signal the operation of a quite different kind of intelligence than is implied by conventional notions of rationality, an intelligence reflecting the properties of the massively parallel computational system that has*

evolved to meet the requirements of existence in a noisy, uncertain, and unstable world. (p.201)

1.2. Hot Hand: Fallacy or Adaptive Strategy

The belief in the hot hand is commonly thought of as the belief that previous successes in an event leads to a greater chance of success in future attempts. Frequently the belief in the hot hand, specifically in basketball performance, is classified as an "illusion" (e.g., Gilovich, Vallone, & Tversky, 1985) and is often included in the collection of erroneous heuristics people use (e.g., gambler's fallacy, availability, representativeness, etc.). Because of this classification, a vast amount of the hot hand literature is concerned with detecting shot dependency/streakiness in real sports data in order to validate/refute people's inclination to perceive the hot hand.

However, too few of these authors are concerned with *why* humans detect this phenomenon, regardless of whether it actually exists (Alter & Oppenheimer, 2006). Even if the belief in the hot hand is not definitively validated in the analysis of sports data, our predisposition to see the world as a sort of "clumping" of resources may in fact be a very useful evolutionary adaption (Wilke & Barrett, 2009). Scheibehenne, Wilke, and Todd (2011) state that:

> *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).

1.2.1. Is Performance Streaky

One of the most prominent evaluations of the persistence of streaks in human performance was Gilovich, Vallone and Tversky's (1985) (referred to as GVT from here on) analysis of basketball shooting. In their analysis of the Philadelphia 76ers and the free throw shooting data for the Boston Celtics (both professional basketball teams in the NBA) they found "no evidence for a positive correlation between the outcomes of successive shots" (p.295). GVT's analysis of a controlled shooting experiment with Cornell's varsity college players also found no statistically significant successive shot dependencies (as measured by conventional statistical test such as serial correlation and Wald-Wolfowitz Runs Test). Without significant evidence for positive shot correlation, they concluded that the hot hand is an illusion and that a player's shooting performance is best predicted by a simple binomial model where the probability of making any given shot is invariant.

Since then, there have been numerous analyses of the validity of the hot hand in sports. For example, a recent meta-analysis by Avugos, Köppen, Czienskowski, Raab, and Bar-Eli (2013) explored 22 different publications with a total of 30 empirical evaluations. They found that a majority of the studies were unable to provide evidence for the hot hand and thus concluded that there is no significant evidence in support of a hot hand effect in sports.

As with many empirical questions, achieving consensus either for or against the existence of the hot hand will never be absolute. However, there are compelling studies that, utilizing proper statistics and empirical methods, shed doubt on the generalized conclusion of Avugos et al. (2013). For example, Bocskocsky, Ezekowitz, and Stein (2014) explored NBA data where they controlled for the difficulty of a shot (since they found players tended to take harder shots following a success) and found significant positive correlation for successful shots. Gilden and Wilson (1995) found that performance of novices in golf putting and darts tended to be streaky (i.e., evidence for the hot hand). Smith (2003) found streaky performance in horseshoe pitching (a sport which has very few confounding variables). Miller and Sanjurjo (2014) found considerable evidence of the hot hand (i.e., streaky performance) in their controlled basketball shooting experiment and Yarri and David (2012) found significant game to game fluctuations in a player's performance in data from the Professional Bowling Association (PBA).

1.2.2. Can People Discriminate Streaky from Random Sequences

Almost 30 years after GVT's renowned hot hand review there is still debate about whether the hot hand exists. But despite the opposition of statisticians, psychologists, etc., who regard the belief in the hot hand as an illusion, why do so many sport fans, coaches, and players still persist in this belief? For example, in a 2010 post game interview, the 11 time NBA champion coach Phil Jackson explained one of his coaching decisions with the comment that you "ride the hot hand" commenting that "even a dumb person would know that" (http://lakersblog.latimes.com/lakersblog/2010/10/phil-jackson-acknowledges-challenge-in-fitting-in-sasha-vujacic-into-rotation.html).

Another example is Raab, Gigerenzer, and Gula's (2012) analysis of the top 26 first-division volleyball players. They found that not only did streaks exist for half of the players, but that the coaches and playmakers (i.e., the player who runs the team's offense) could in fact detect this variability in a player's performance and apply this knowledge to make strategic defensive/offensive decisions. Cooper (2013) also found that in four different empirical evaluations with simulated data, participants were able to discriminate a streaky sequence (non-random process) from a Bernoulli random process significantly better than chance.

1.2.3. Does Context Matter

The effect of context in sports performance is another important aspect to consider. GVT suggested that not only do streaks not exist in basketball player's performance, but that a player's shooting is analogous to the flips of a coin. Drawing this parallel is to say that successive trials (e.g., flips or shots) are independent, the probability of success remains the same for any given trial (i.e., stationary probability), and that *context does not matter*. It is fair to say that coin flips would not be influenced by having been flipped in Los Angeles, California, versus Denver, Colorado. However, can the same be said for a player shooting in Los Angeles (their team hometown) versus Denver (their opponent team's hometown)?

1.3. Structure of Current Research

In order to further explore the idea of the hot hand and people's streaky/pattern perception, a series of Monte Carlo simulations were generated in order to represent various types of plausible constraints that could impact a basketball players shooting performance. Once the binary sequences were generated, it made it possible to explore not only the statistical discriminability of these sequences, but also the human discriminability. An empirical experiment was then conducted using the same generating rules as the Monte Carlo simulations, where the probability parameters of the sequences were calibrated to reflect actual shooting percentages of NBA players, in order to explore human discriminability of Bernoulli random sequences from alternative performance dependent learning constrained sequences.

A primary goal of this research was to explore whether humans can make the discrimination between Bernoulli random and performance dependent learning constrained sequences that the statistical tests were not able to discriminate. Furthermore, by looking at the level of success (i.e., variance in the number of points scored) of NBA point guards, insights into whether context indeed matters with regards to performance during home and away games was investigated.

2. MONTE CARLO SIMULATIONS

2.1. Simulation Method

Binary shooting sequences were generated using Monte Carlo simulations in order to represent different types of real-world constraints that could impact a players shooting performance. The criteria that guided the different types of constraints were based on the dependencies of trials coupled with the stationary properties of the rules governing the sequence, 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 (non-stationary).

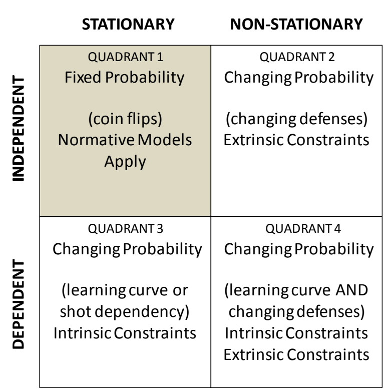


Figure 1. This diagram illustrates four types of processes as a function of whether the generating rules are independent and stationary. The constraints associated with Quadrants 2, 3, and 4 were chosen to reflect plausible constraints on sports performance.

2.1.1. Quadrant 1

The binary sequences generated from quadrant 1 represent independent, stationary processes. These Bernoulli sequences correspond to an unconstrained process (i.e., random process), where the fixed probabilities of success for each of the three simulations are 0.3, 0.5, and 0.8.

2.1.2. Quadrant 2

These simulations represent a sequence being sampled from two different Bernoulli processes (i.e., an independent, non-stationary process). One process had a probability of success of 0.6 (hot streak) and the other had a probability of success of 0.2 (cold streak). The rate of alternation between these two Bernoulli processes was representative of a possible scenario where a player had different success rates against a more or less aggressive defense. Three simulations were generated with the probability of switching equal to 0.10, 0.25, and 0.50. The process with the lowest probability of switching (.1) had the streakiest sequences (i.e., less switching means longer hot and cold streaks).

2.1.3. Quadrant 3

This quadrant contains three different conditions governed by two separate rules that represented a dependent, stationary process. The first condition had trials based on previous local dependencies. In this shot (trial) dependent condition two different sequence types were generated. The first sequence type was based on the success of the previous trial. If the previous trial was a success the probability of making the next trial became 0.7 and if it were a miss the probability became 0.3. The second sequence type in the shot dependency condition was determined by the successes of the preceding five trials: if 2 or fewer successes, p(success) = 0.3; if 3 successes, p(success) = 0.5; if 4 or more successes, p(success) = 0.7.

The third condition in Quadrant 3 had trials based on previous global dependencies. The probability of success was produced exponentially in relation to the number of shots taken (i.e., trials produced). This dependent, stationary process is representative of a player's increasing success rate with more practice (i.e., simulated learning curve). All simulations in this condition had an initial probability of success equal to zero and an asymptotic probability of success equal to 0.8. The three simulations had either a learning rate (k) of .001, .003, or .005.

2.1.4. Quadrant 4

In Quadrant 4 there was only one simulation generated. This process combined the exponential learning rate used in Quadrant 3 (initial success probability = .2; asymptotic success probability = .6; learning rate = .002) and the alternating probability of success used in Quadrant 2. Therefore, the actual probability was increased by .1 above or below the probability computed from the learning function. As with Quadrant 2 the alteration between +.1 or -.1 was based on a fixed probability equal to .1. This resulted in persistent hot (+.1) or cold (-.1) streaks on top of a typical learning curve.

2.2. Simulation Analysis

To determine whether the simulations led to sequences that are statistically different than expected for a "random" or Bernoulli process two tests were used, the Wald-Wolfowitz Runs Test and the slope of the frequency response function (Beta) resulting from Fourier Analysis. 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. Figure 2 through Figure 6 show an example of the spectral pattern for log (power) x log (frequency) for one of the conditions in each of the 4 quadrants shown in Figure 1.

For each of the simulation conditions ten binary sequences were generated, each with 1024 data points (trials). This yielded ten beta values for each of the simulation conditions. 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 one-tailed t-tests were used to evaluate whether the slopes in these simulations were reliably less than zero. One of the main research questions was to see which conditions showed deviations from the expectations of a Bernoulli process for the runs test: large absolute z-score and/or slopes that are significantly different than zero for the spectral analysis.

In order to see if any of the simulation conditions deviated from the expectations of a Bernoulli process as a function of sample size, each condition was also analyzed using different sample sizes. The 1024 data point sequences along with the first 512 data points in those sequences were analyzed using both the Runs Test and spectral analysis. Furthermore, the first 800 data points from the 1024 data point sequences were also analyzed using the Runs Test.

2.3. Simulation Results

Table 1 summarizes the results for the twelve simulations that were conducted. The table includes the overall mean performance (percent success), the results of the Wald-Wolfowitz Runs Tests using various sample sizes, and the mean slopes (betas) and one-tailed t-test results from the spectral analysis using various sample sizes. It was anticipated that none of the conditions in Quadrant 1 nor the 50% chance of alternation condition in Quadrant 2 would show deviations from the expectations of a Bernoulli random process for either the runs test (small absolute z-score) or for the spectral analysis (slopes that are not significantly different from zero), across all sample sizes. The p(hit) = 0.8 condition in Quadrant 1 was the only condition 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.

For the remaining eight conditions spanning Quadrants 2-4, six of the conditions showed deviation from the expectations of a Bernoulli process across all sample sizes for both statistical tests used. However, both the 25% chance of alternation condition in Quadrant 2 and the simple learning curve condition with k = .001 in Quadrant 3 did *not* show deviations from the expectations of a Bernoulli process, but only when 512 data points were used for the Runs Test. These results confirm that with adequate sample sizes (at least 800 samples for the Wald-Wolfowitz test and 1024 samples for the frequency-based analysis) all but one of the constraints could be detected using these statistical methods. The exception was the condition where the probability alternated with .50 probability between hot (.6) and cold (.2).

	N = 1024		Runs Test N = 512	Runs Test N = 800	Runs Test N = 1024	Frequency Analysis Beta Slope N = 512			Frequency Analysis Beta Slope N = 1024		
	Mean	SD	|Z|	|Z|	|Z|	Mean	SD	*t*	Mean	SD	*t*
Quadrant 1											
Bernoulli Processes											
p(hit) = 0.3	0.306	0.013	0.04	0.18	0.44	0.02	0.06	0.87	0.02	0.04	1.50
p(hit) = 0.5	0.500	0.018	0.26	0.06	0.02	0.03	0.05	1.94	0.01	0.06	0.26
p(hit) = 0.8	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) = 0.2 and p(hit) = 0.6											
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 *
Simple Learning Curves											
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 *
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. * $p < .05$ for two-tailed z-test for runs; * $p < .05$ for one-tailed t-test for slope (beta). N = 512 and N = 800 indicate that the beginning 512 and 800 (respectively) data points from the 1024 data point sequence were used.

Table 1. The overall mean performance (percent success), the results of the Wald-Wolfowitz Runs Tests across sample sizes, and the mean slopes (betas) and one-tailed t-test results from the spectral analysis across sample sizes.

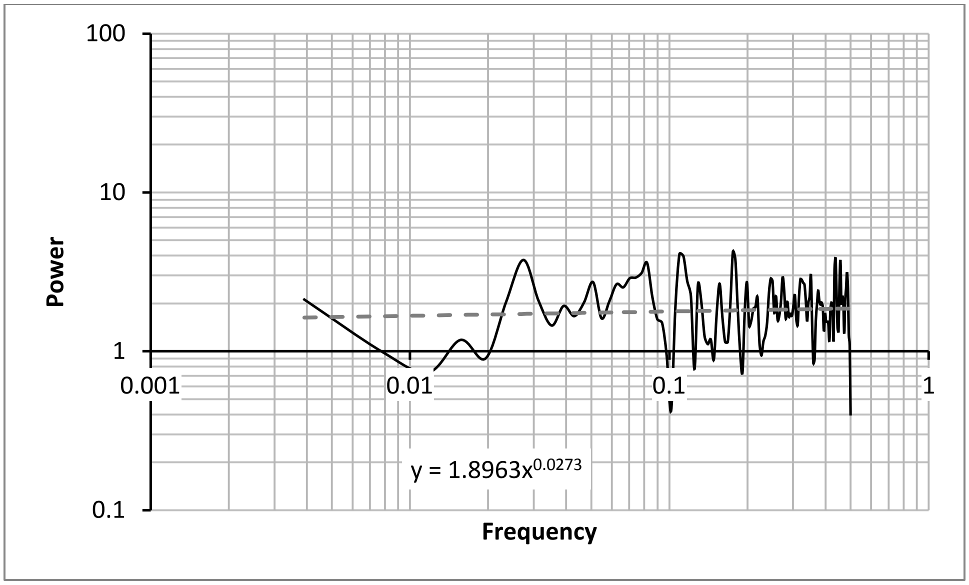


Figure 2. A sample of the spectral pattern for log (power) x log (frequency) for one of the three conditions in Quadrant 1 in which p(hit) = 0.3.

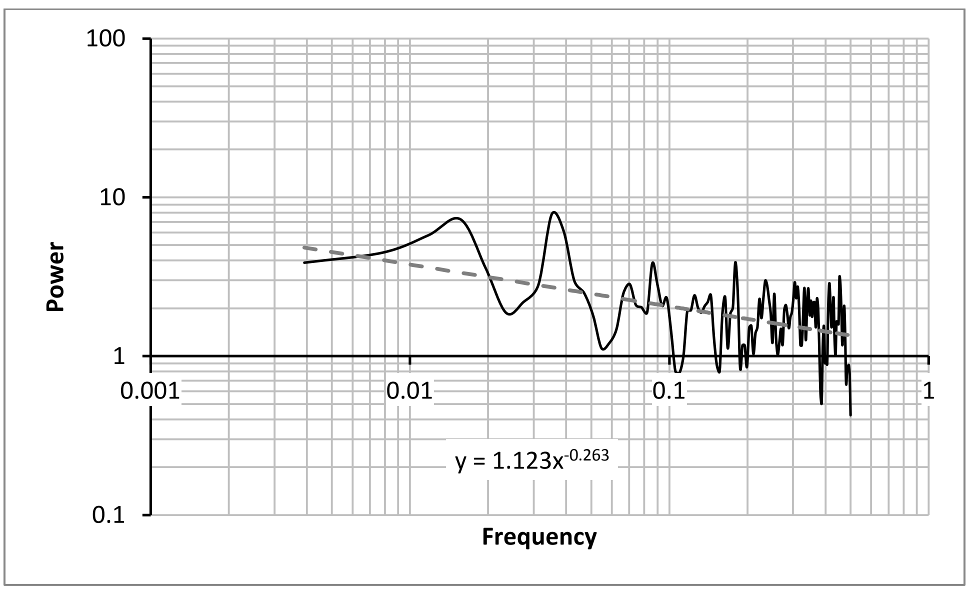


Figure 3. A sample of the spectral pattern for log (power) x log (frequency) for one of the three conditions in Quadrant 2 in which the probability of success varied between p(hit) = 0.2 and p(hit) = 0.6 with a probability of alternating success rates of .10.

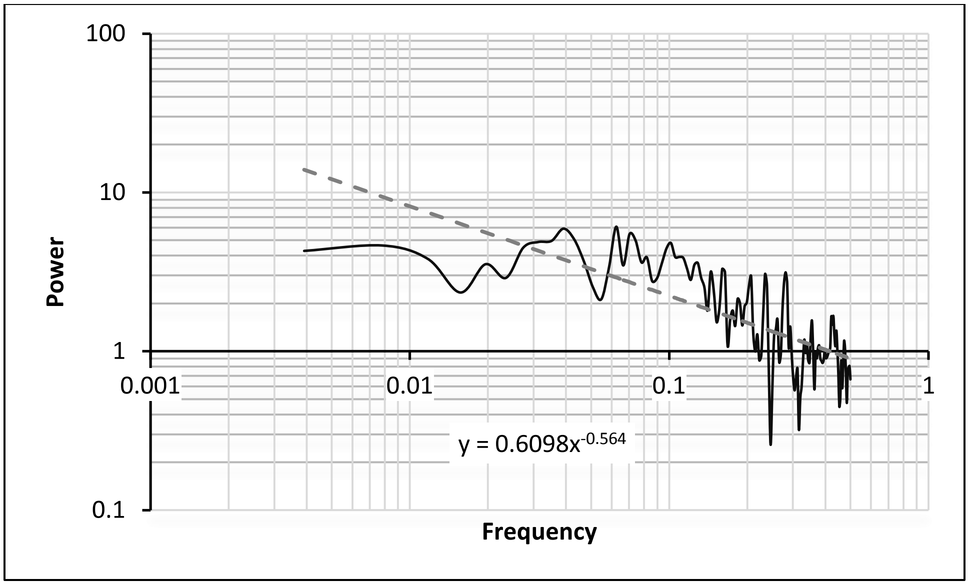


Figure 4. A sample of the spectral pattern for log (power) x log (frequency) for one of the five conditions in Quadrant 3 where there was a 1 shot dependency and the probability of success varied between p(hit) = 0.3 if the previous shot was missed and p(hit) = 0.7 is the previous shot was made.

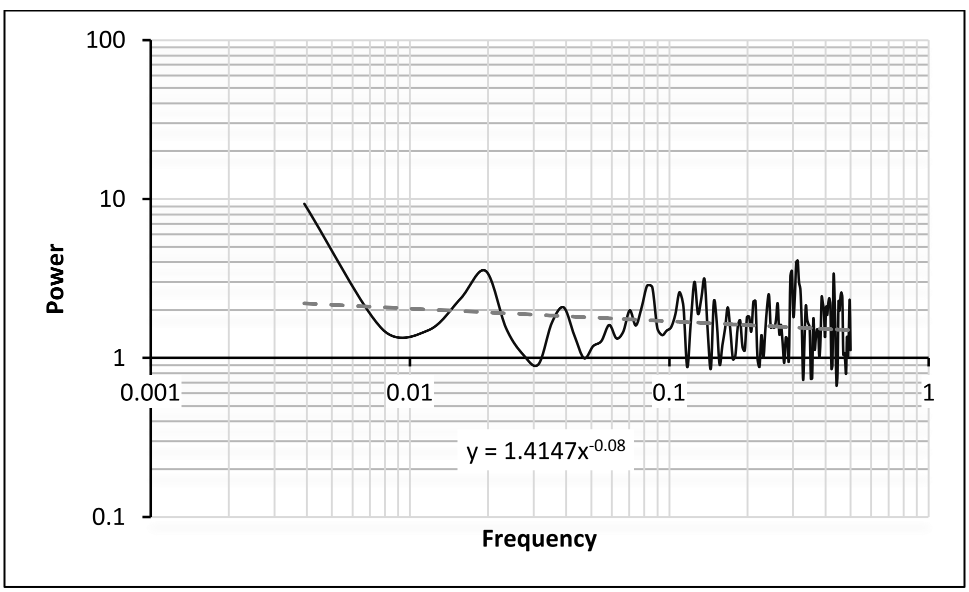


Figure 5. A sample of the spectral pattern for log (power) x log (frequency) for one of the five conditions in Quadrant 3 where there was a simple learning curve with k = .001, an initial p(hit) = 0.2, and an asymptotic p(hit) = 0.6.

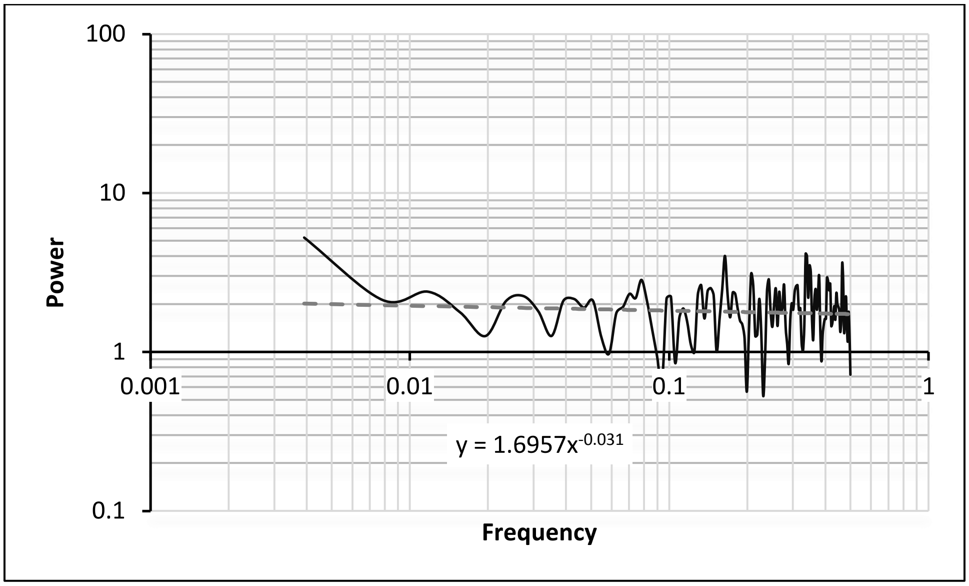


Figure 6. A sample of the spectral pattern for log (power) x log (frequency) for the condition in Quadrant 4 where there was a simple learning curve with k = .002, an initial p(hit) = 0.2, an asymptotic p(hit) = 0.6, and an alternation rate of switching 10 percent of the time between either .1 above or .1 below the computed probability from the learning function.

3. MONTE CARLO SIMULATIONS OF A PERFORMANCE DEPENDENT LEARNING CURVE

3.1. Introduction

Cooper (2013) conducted five empirical experiments that used very similar or identical data as that generated in the previously described Monte Carlo simulations in Table 1. These experiments were done in order to test human discriminability of constrained versus randomly generated sequences. The first experiment had participants discriminate between two Bernoulli processes, while the other four experiments asked them to discriminate between a Bernoulli process and an alternative one, which spanned quadrants 2-4 in Figure 1. For all five experiments participants could accurately discriminate between the types of processes significantly better than chance.

One plausible constraint that had yet to be explored however was a sequence generated from a performance (success) dependent learning curve. This type of constraint represents a basketball player who progressively becomes a better shooter (i.e., makes more shots) as the number of shots they have *successfully* made over the course of their "season" increases. This type of constraint thus falls in Quadrant 3 since the trials would be dependent and the governing rule stationary.

The plausibility of this constraint in a real world setting prompted an exploration of a performance dependent learning constraint, which began with generating Monte Carlo simulations. Not only was the statistical discriminability of this constraint explored, but also the human discriminability. The same performance dependent learning sequences generated for the Monte Carlo simulations were used not only in statistical analyses to see if the sequences deviated from the expectations of a Bernoulli random process, but also in an empirical discrimination task to see if people were able to make the discrimination between a Bernoulli generated sequence and a sequence based on the performance dependent learning constraint.

3.2. Simulation Method

The simulations were implemented and analyzed using Microsoft Excel software and MATLAB. The first step in the simulation was to generate binary strings (1s/successes, and 0s/misses) that adjusted the probability of success in an exponential fashion as a function of the number of the successful (i.e., made) shots. Different simulations were generated for the different learning rates of .001, .003, and .005. Using these learning rates, two different initial and asymptotic probabilities of success were used. The first had the initial probability of success equal to .33 and asymptotic probability of success equal to .61, and the second set had an initial of .20 and asymptotic of .80. The moving probability of success is computed using the following formula, where I = initial p(hit), A = asymptotic p(hit), k = learning rate, S = sum of all prior hits:

$$\mathrm{p(hit)} = \mathrm{I} + ((\mathrm{A} - \mathrm{I}) \times (1 - \mathrm{e}^{(-(\mathrm{k}) \times \mathrm{S})}))$$

The initial and asymptotic probabilities of .33 and .61 were determined by taking the NBA players with the lowest and highest (respectively) percentage of field goals made for a given season (averaged across seven regular seasons). The player with the lowest/highest field goal percentages had to have taken at least 246 shots (i.e., on average 3 shots per game in an 82 game season) in that season, thus excluding players who have had non-representative shooting performance. The success rate of the Bernoulli process was computed by taking the average of the median players' field goal percentage (using the same 246 shot minimum per season criterion) from each of the same seven seasons used to find the initial and asymptotic probabilities for the performance dependent learning curve. The initial and asymptotic probabilities of .20 and .80 were used in order to keep the parameters of this simulation similar to that used in the simple learning curve simulation (shown in Table 1).

3.3. Simulation Analysis

In order to determine whether the performance dependent learning curve simulations lead to sequences that are statistically different than expected for a random or Bernoulli process, the same two tests (Runs Test and frequency analysis) and statistical methods were used as those previously described for the simulations in Table 1, along with Bayesian model comparisons (described in Recoding Method). Figures 7 and 8 show an example of the spectral pattern for log (power) x log (frequency) for one of the conditions using the initial and asymptotic probabilities of .33 and .61 and for one of the conditions using the initial and asymptotic probabilities of .20 and .80.

In order to see when and which simulation conditions deviated from the expectations of a Bernoulli process, the same exploration of the effects of sample size was also performed for these simulations as that described for the simulations in Table 1. However, in addition to the sample sizes used for the simulations in Table 1, the runs test analyses were also performed using the beginning, middle, and end 800 data points of the 1024 data point sequence. This was done in order to explore the results of the statistical tests when there is an exponential increase in the number of successes, which is the case for the performance dependent learning curve conditions.

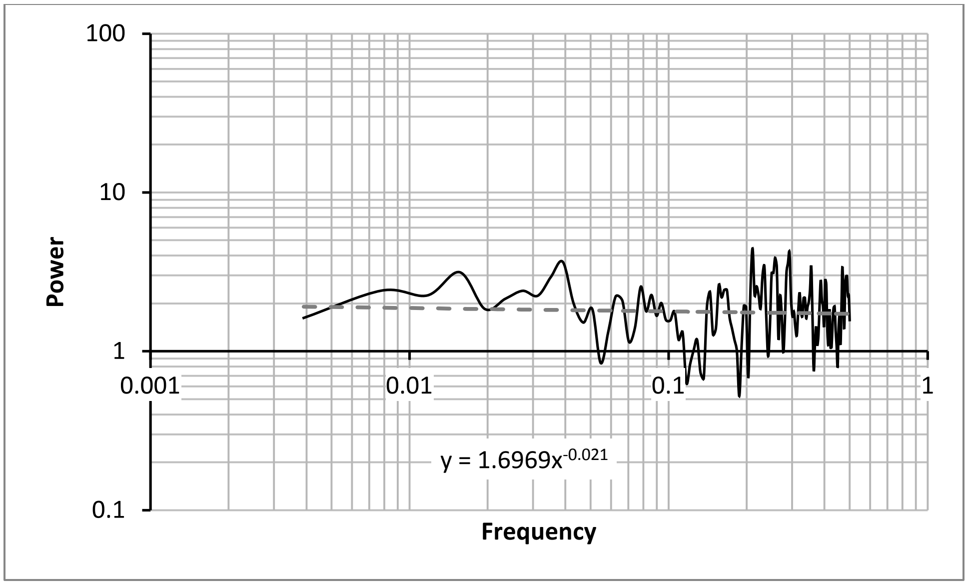


Figure 7. A sample of the spectral pattern for log (power) x log (frequency) for the performance dependent learning condition with k = .001, an initial p(hit) = 0.20, and an asymptotic p(hit) = 0.80.

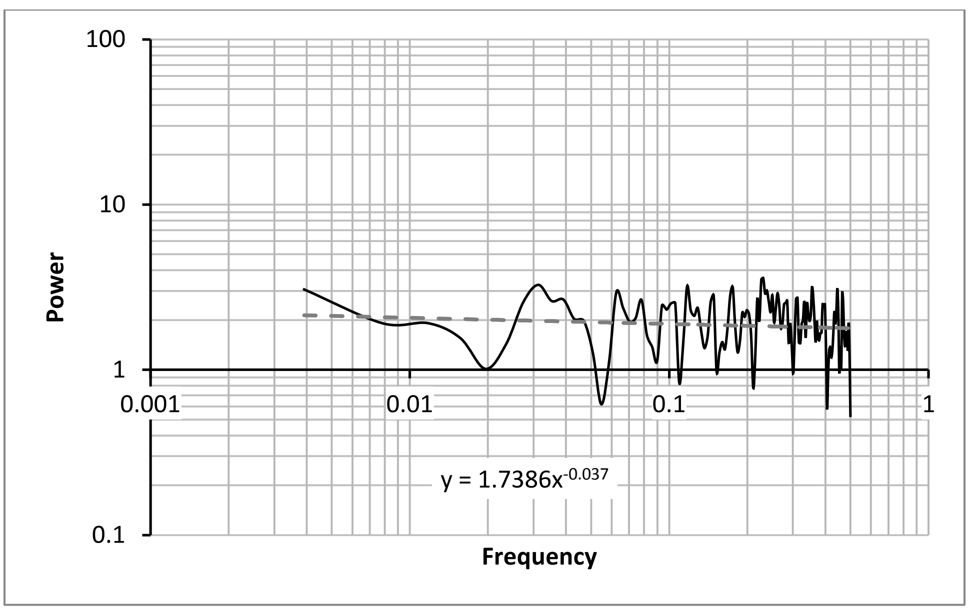


Figure 8. A sample of the spectral pattern for log (power) x log (frequency) for the performance dependent learning condition with k = .001, an initial p(hit) = 0.33, and an asymptotic p(hit) = 0.61.

3.4. Simulation Results

Table 2 summarizes the results for the twelve simulations that were conducted. The table includes the overall mean performance (percent success), the results of the Wald-Wolfowitz Runs Tests using various sample sizes, and the mean slopes (betas) and one-tailed t-test results from the spectral analysis using various sample sizes. For the three performance dependent learning curve simulations using the initial and asymptotic probabilities of .33 and .61 (respectively), none deviated from the expectations of a Bernoulli process for the runs test (small absolute z-score). There was only one deviation from the expectations for a Bernoulli process with the spectral analysis (slopes that are not significantly different from zero) and that was with the learning rate of .005 and a small sample size (512). Thus, this constraint was not detected by the conventional statistical tests.

Of particular interest is the performance dependent learning curve condition with k = .005 and the initial and asymptotic probabilities of .33 and .61, since this condition was used in the empirical experiment (using the beginning 800 data points). Recall that the probability parameters for this condition were calibrated to reflect actual shooting percentages of NBA players. For this condition, only using the weakest sample size of 512 data points resulted in a significant deviation from the expectations of a Bernoulli process; the rest of the analyses in this condition did not significantly deviate from the expectations of Bernoulli random process, meaning that using the beginning 800 data point the statistical tests could *not* make the discrimination.

For the three simulations using the initial and asymptotic probabilities of .20 and .80, deviation from the expectations of a Bernoulli process occurs with the learning rate of k = .003 and k = .005. With k = .003, the only simulation to show deviation from the expectations of a Bernoulli process occurs with the Runs Test using the strongest sample size of 1024 data points. With k = .005, the only simulation to *not* show deviation from the expectations of a Bernoulli process occurs with the Runs Test using 512 data points and the end 800 data points.

				Runs Test					Frequency Analysis Beta Slope			Frequency Analysis Beta Slope		
		N = 1024		N = B.512	N = B.800	N = M.800	N = E.800	N = 1024	N = 512			N = 1024		
		Mean	SD	|Z|	|Z|	|Z|	|Z|	|Z|	Mean	SD	*t*	Mean	SD	*t*
Quadrant 3														
Performance Dependent Learning Curves														
p(hit) = 0.33 and p(hit) = 0.61	k = 0.001	0.380	0.019	0.19	0.16	0.48	0.51	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	0.48	0.30	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**	**0.24**	**0.48**	**0.75**	**-0.05**	**0.06**	**-2.37 ***	**-0.02**	**0.05**	**-1.64**
p(hit) = 0.20 and p(hit) = 0.80	k = 0.001	0.273	0.013	0.17	0.16	0.08	0.05	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	0.92	1.07	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 *	2.56 *	1.81	4.09 *	-0.11	0.06	-5.63 *	-0.09	0.05	-5.53 *

*Note. * p < .05 for two-tailed z-test for Runs Test; * p < .05 for one-tailed t-test for slope (beta). B.512 and B.800 indicates the beginning 512 and 800 data points, M.800 indicates the middle 800 data points, and E.800 indicates the end 800 data points from the 1024 data point sequence. Empirical experiment testing human discriminability used trials generated with initial p(hit) = 0.33, asymptotic p(hit) = 0.61, and k = .005.*

Table 2. The overall mean performance (percent success), the results of the Wald-Wolfowitz Runs Tests across sample sizes, and the mean slopes (betas) and one-tailed t-test results from the spectral analysis across sample sizes.

4. EMPIRICAL STUDY METHOD

4.1. Introduction

As previously mentioned, Cooper (2013) conducted five empirical experiments based on the previously described Monte Carlo simulations in Table 1 in order to test human discriminability of constrained versus randomly generated sequences. A string of 800 events was presented to the participants that represented a binary string of basketball shots (i.e., successes and misses). The first experiment had participants discriminate between two Bernoulli processes, while the other four experiments asked them to discriminate between a Bernoulli process and an alternative one, which spanned quadrants 2-4 in Figure 1. These experiments used the same governing rules and parameters for generating data as that described for the previous Monte Carlo simulations. For all five experiments participants could accurately discriminate between the types of processes significantly better than chance. One plausible constraint that had not been empirically explored however was performance (success) dependent learning. After finding that the statistical tests were *not* able to detect the performance dependent learning constraint as deviating from a Bernoulli random process, the question still remained whether or not humans could make this discrimination.

4.2. Participants

As with Cooper's (2013) evaluation of human discrimination, 16 participants were used in this experiment. 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.

4.3. Equipment

The experiment was carried out using an Apple iMac. The program used was developed through the Java runtime environment for Cooper's (2013) experiments. A standard optical mouse was used to control the display and select the desired options.

The interface included a window through which the participant could view the binary sequence. Three different window sizes were used where either 1, 4, or 16 events were visible in a window. There were a total of 800 events in each sequence available to the participants. 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 window over the sequence. A number was displayed above the slider to indicate the number for the first visible event in the window. After viewing 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 numbers of correct and wrong answers. This updated immediately following each response. A screen shot of this interface with the 16 event visible window size is given in Appendix A.

4.4. Procedure

Participants were allowed six practice trials where they saw one trial from the constrained process and one trial from the control (Bernoulli random) process for each viewing window size. 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, they 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. The instructions to participants were designed to provide some insight into the specific nature of the task, based on Nickerson's (2002) recommendation for clear, non-ambiguous, and precise instructions for human production and perception of randomness. The participant instructions are provided in Appendix B.

At the beginning of each trial the window was positioned in the middle of the 800-event sequence. For each trial participants were allowed to freely explore the sequence 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 time limited. After completing 30 trials the participants were asked question regarding their strategy, level of engagement, and performance (see Appendix C for complete participant debriefing). After the post block questions participants then proceeded to the next block. The time needed to complete the experiment ranged from 50-140 minutes (average of 86 minutes). In contrast, Cooper's (2013) participants' completion time for a very similar discrimination task ranged from 20-45 minutes (refer to the Discussion of the Empirical Experiment for further details).

4.5. Design

One experiment was conducted in which participants were presented with sequences representing binary strings of basketball shots (successes and misses). In the experiment participants were presented with either a sequence generated by a Bernoulli (control) process or an alternative (constrained) process. The participants were asked to discriminate between "steady" shooters and "improving" shooters. The response button labeled "streaky" was used to indicate the improving shooter and the response button labeled "steady" was used to represent the steady shooter. The steady shooter had a constant p(hit) = 0.44 and the improving shooter's performance was governed by the performance dependent learning curve with a learning rate = .005, a starting p(success) = 0.33, and an asymptote at p(success) = 0.61 (recall these parameters were based on actual NBA shooting performance). This was identical to the condition generated for the performance dependent learning Monte Carlo simulation (using the beginning 800 data points), which was *not* detected by the Runs Test or spectral analysis as deviating from the expectations of a Bernoulli random process.

4.6. Independent Variables

The experiment involved 3 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 sequences generated by the control process and 15 presentations of sequences generated by the constrained process, presented in a the same random order for all participants.

4.7. Dependent Variables

The dependent measures include the proportion of correct responses (categorized in a signal detection matrix as proposed by Green and Swets, 1966), an estimated *d'* for each subject, a response bias value for each subject, 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 subject 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.

5. EMPIRICAL RESULTS

5.1. Measuring Sensitivity (*d'*)

Recall that the Runs Test and spectral analysis did *not* detect the performance dependent learning constraint (with k = .005, initial p(hit) = .33 and asymptotic p(hit) = .61) as being significantly different from a Bernoulli random process. However, the analysis of human performance showed the distributions of adjusted d primes across all window sizes to have no overlap with *d'* = 0, indicating that humans *were* able to make the discrimination that the statistical tests could not (see Table 3).

For Window Size 16 the upper and lower bounds of the 97.5% confidence interval for *d'* were 1.15 and 0.50 respectively. For Window Size 4 the upper and lower bounds of the 97.5% confidence interval for *d'* were 1.27 and 0.65. And for Window Size 1 the upper and lower bounds of the 97.5% confidence interval for *d'* were 0.96 and 0.44. These results indicate that participants were able to discriminate the Bernoulli generated sequence from the alternative constrained sequence reliably better than chance for each window size, although their performance was best in Window Size 4 and worst in Window Size 1.

Pairwise comparisons of the *d'* distributions as a function of window size showed the posterior probability that group *d'* for Window Size 4 is larger than Window Size 1 is 0.89. The posterior probability that the group *d'* for Window Size 4 is larger than Window Size 16 is 0.74. And the posterior probability that the group *d'* for Window Size 16 is larger than Window Size 1 is 0.71. These results indicate that there was not strong evidence for a difference in d prime between the window sizes; to the extent that there is a difference, the trend is toward better performance in Window Size 4 and worse performance in Window Size 1.

	Hit Rate		False Alarm Rate		Adjusted d'		Count	Bias Value	
	Mean	SD	Mean	SD	Mean	SD		Mean	SD
Window = 16	0.64	0.03	0.33	0.05	0.82	0.17	12	0.04	0.09
Window = 4	0.67	0.03	0.31	0.05	0.96	0.16	15	0.03	0.08
Window = 1	0.61	0.04	0.34	0.03	0.70	0.13	15	0.06	0.07

Note. Mean and Standard Deviation (SD) were taken across participants.Count refers to the number of participants (out of 16) with an adjusted d' reliably different from 0.

Table 3. Mean percentage rate of hits and false alarms, adjusted *d'*, and response bias *c* value as a function of window size. All of the posterior means are from the Bayesian estimates.

5.2. Response Bias *c*

Due to our evolutionary predisposition to see clumping or patterns of resources (Wilke & Barrett, 2009), a reasonable place to explore human strategy was through the participant response bias *c* in this discrimination task. It was anticipated that perhaps participants would select the streaky option more often than the steady option (judge sequences as being constrained more often than random) based on the idea that people have a predisposition to see patterns. The following formula was used for computing individual participants response bias *c*, where z(H) is the z transform of the hit rate and z(F) is the z transform of the false alarm rate (Macmillan & Creelman, 1991):

$$c = -\frac{[\mathrm{z(H)} + \mathrm{z(F)}]}{2}$$

The distribution of response bias values across all window sizes did overlap with *c* = 0, indicating that there was not a reliably different response bias across participants. For Window Size 16 the upper and lower bounds of the 97.5% confidence interval for response bias *c* values were 0.22 and -0.11 respectively; a positive response bias value indicates a bias for selecting the steady (Bernoulli random) option and a negative value indicates a bias for selecting the streaky (constrained) option. For Window Size 4 the upper and lower bounds of the 97.5% confidence interval for response bias were 0.19 and -0.13. And for Window Size 1 the upper and lower bounds of the 97.5% confidence interval for response bias were 0.19 and -0.07. These results do not indicate strong evidence that participants had a response bias for selecting either the streaky or steady option when analyzed across participants for each window size.

Pairwise comparisons of the bias distributions as a function of window size showed that the posterior probability that the response bias in Window Size 1 is larger than Window Size 4 is 0.65. The posterior probability that the response bias for Window Size 1 is larger than Window Size 16 is 0.59. And the posterior probability that the response bias for Window Size 16 is larger than Window Size 4 is 0.53. These results indicate that there was not strong evidence for a difference in response bias *c* between the window sizes.

5.3. Individual Participants

After plotting each participant's *d'* and bias value for each window size, a strong linear relationship was found for Window Sizes 16 and 4 (see Figures 9 and 10). More specifically, participants who did better at the discrimination task with Window Size 16 and 4 showed a stronger bias to select the steady option (Bernoulli random). The correlation between a participants *d'* and bias value for Window Size 16 was .99 and .96 for Window Size 4, which are almost perfect correlations between participants *d'* and bias value. For Window Size 1 however, the relationship between participant *d'* and bias values was substantially weaker. Only one person showed a bias to respond streaky in this window size and there was a correlation of -.51 between participant *d'* and response bias (see Figure 11). Further commentary on these results can be found in the Empirical Experiment segment of the Discussion section.

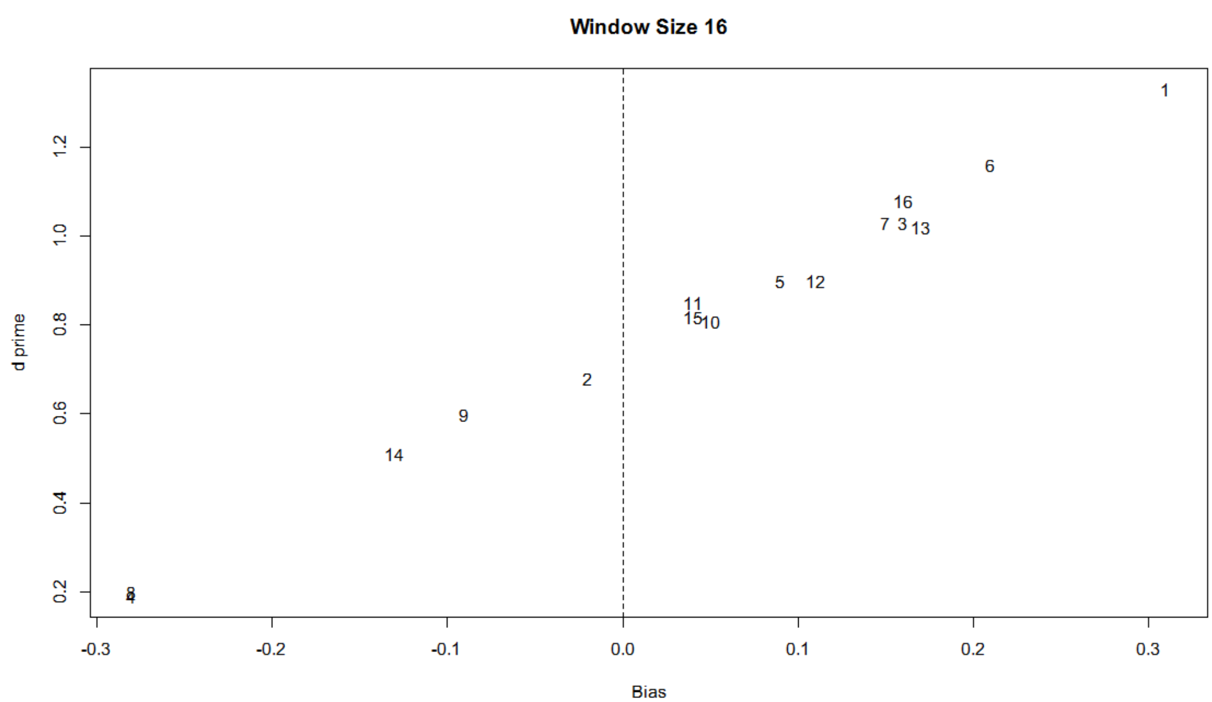


Figure 9. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 16. Each plotted number indicates a participant.

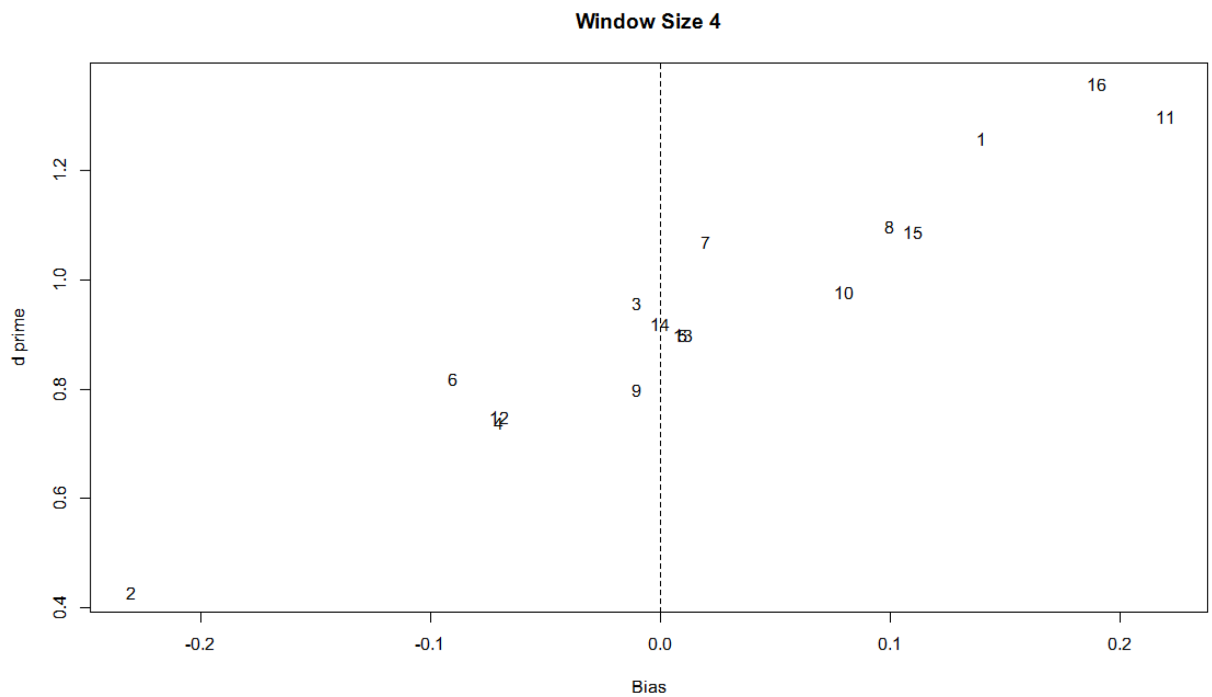


Figure 10. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 4. Each plotted number indicates a participant.

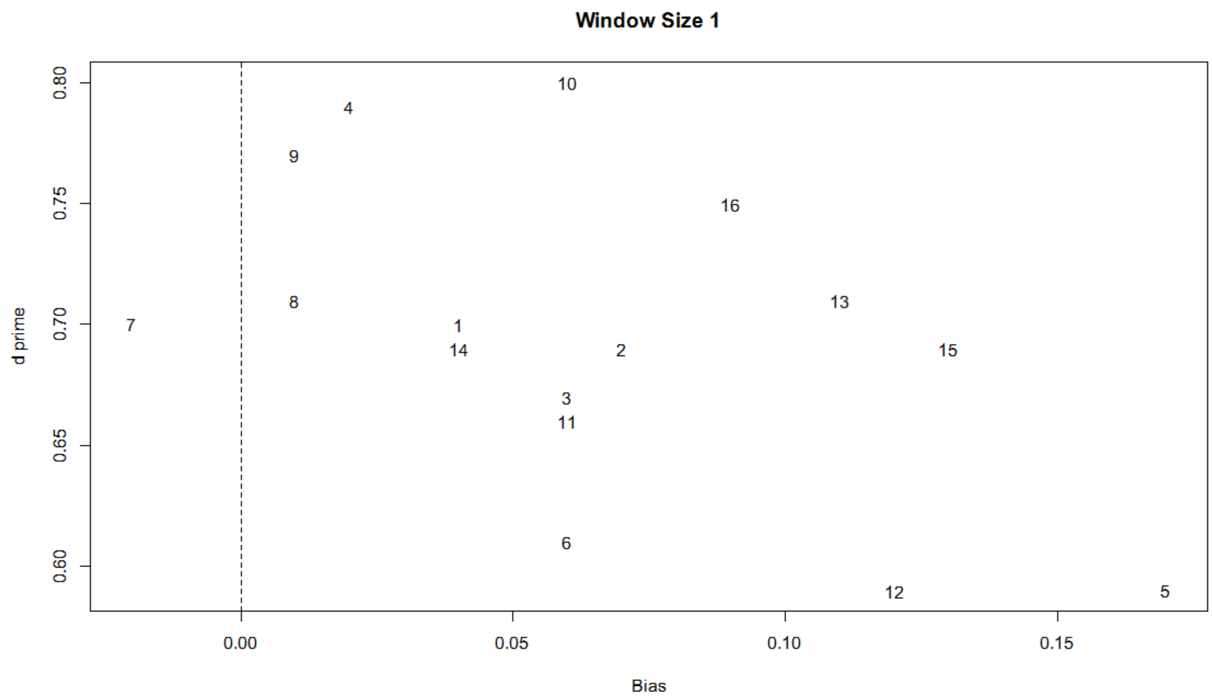


Figure 11. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 1. Each plotted number indicates a participant.

5.3.1. Participant Strategies

There were two general strategies participants used to complete this discrimination task; either they would look throughout the 800 events in the sequence, or they would focus primarily on the beginning and/or end events in the sequence. Furthermore, some participants would switch or alter their strategy over the course of the experiment, while others would stick with the same strategy. The vast majority of participants who were switching strategies were doing so with the hopes of performing the task more accurately and/or efficiently.

A particularly interesting finding was that *none* of the top three participants (based on *d'*) for Window Size 16 or Window Size 4 were in the top three performers for Window Size 1. This finding indicates that perhaps strategies that worked well for

Window Sizes 16 and 4 did not continue to work as well for Window Size 1. An additional finding was that only in Window Size 1 did people mention using a strategy that deliberately leveraged visual processing. More specifically, participants would scroll through the events so quickly that they would induced a motion smearing effect out of the orange (made shot) and grey (missed shot) icons (see Appendix A for image of static interface). So rather than cognitively averaging the hits and misses, these participants were visually averaging them based on hue. This finding is of particular importance because it provides an excellent example of an ecologically valid and adaptive strategy. In an actual NBA game this motion smearing strategy is not even applicable. However, when this same discrimination task is set up differently, it affords an alternative range of possible solutions to the problem.

6. EMPIRICAL RESULTS RECODED

6.1. Recoding Method

With known models of the processes underlying the sequences being generated, Bayesian model comparisons were conducted for all of the empirical experiment trials (N = 90) post hoc. Both models represented optimal performance assuming full knowledge of the task parameters. One model assumed a random Bernoulli process where the p(hit) = .44 (a “steady” shooter), and the other model assumed the performance 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:

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

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 as being a steady sequence and vise 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 performance dependent learning curve sequence (streaky) in Window Size 16 and two trials were recoded from performance dependent learning curve sequences to Bernoulli generated sequences, one in Window Size 16 and one in Window Size 4.

6.2. Recoded Results

6.2.1. Recoded Measure of Sensitivity (*d'*)

The posterior probability that the overall d prime across all window sizes after recoding is larger than before recoding is 0.64 and the correlation between participant d primes before and after recoding is 0.95, which is not strong evidence for a difference in d prime after the three experimental trials were recoded. The distributions of adjusted d primes across all window sizes showed no overlap with *d'* = 0 after the three experimental trials were recoded (see Table 4), which still indicates that humans *are* able to make the discrimination between Bernoulli random sequences and performance dependent learning constraint sequences, while the statistical tests could not.

For Window Size 16 the upper and lower bounds of the 97.5% confidence interval for *d'* were 1.15 and 0.51 respectively. For Window Size 4 the upper and lower bounds of the 97.5% confidence interval for *d'* were 1.46 and 0.88. And for Window Size 1 the upper and lower bounds of the 97.5% confidence interval for *d'* were 0.96 and 0.46. These results indicate that participants were still able to discriminate the Bernoulli generated sequences from the alternative constrained sequences reliably better than chance for each window sizes after the three experimental trials were recoded.

Pairwise comparisons of the *d'* distributions as a function of window size after the three experimental trials were recoded showed that the posterior probability that *d'* for Window Size 4 is larger than Window Size 1 is 0.99. The posterior probability that *d'* for Window Size 4 is larger than Window Size 16 is 0.93, and the posterior probability that *d'* for Window Size 16 is larger than Window Size 1 is 0.78. These results indicate that there was strong evidence for a difference in *d'* between Window Size 4 and Window Size 1 and marginally strong evidence for a difference in *d'* between Window Size 4 and

Window Size 16. There was not strong evidence for a difference between Window Size 16 and Window Size 1 however. To the extent that there is a difference, the trend is toward better performance in Window Size 4 and worse performance in Window Size 1 after the trials were recoded.

	Hit Rate		False Alarm Rate		Adjusted d'		Count	Bias Value	
	Mean	SD	Mean	SD	Mean	SD		Mean	SD
Window = 16	0.65	0.03	0.32	0.05	0.86	0.17	12	0.04	0.08
Window = 4	0.72	0.03	0.29	0.04	1.16	0.14	16	-0.02	0.07
Window = 1	0.61	0.03	0.34	0.03	0.70	0.12	15	0.06	0.06

Note. Analysis after the recoding of trials. Mean and Standard Deviation (SD) were taken across participants. Count refers to the number of participants (out of 16) with an adjusted d' reliably different from 0.

Table 4. Mean percentage rate of hits and false alarms, adjusted *d'*, and response bias *c* value as a function of window size after recoding three experimental trials.

6.2.2. Recoded Response Bias *c*

The posterior probability that the response bias value across all window sizes before recoding is larger than after recoding is 0.55 and the correlation between participant bias values before and after recoding is 0.96, which is not strong evidence for a difference in response bias *c* after the three experimental trials were recoded. The distributions of response bias values across all window sizes after recoding the three experimental trials still overlap with $c = 0$, indicating that participants still did not have a strong response bias.

For Window Size 16 the upper and lower bounds of the 97.5% confidence interval for response bias were 0.22 and -0.11 respectively; as with before recoding, a positive response bias value indicates a bias for selecting the steady (Bernoulli random) option and a negative value indicates a bias for selecting the streaky (constrained) option. For

Window Size 4 the upper and lower bounds of the 97.5% confidence interval for response bias were 0.13 and -0.17. And for Window Size 1 the upper and lower bounds of the 97.5% confidence interval for response bias were 0.18 and -0.06. As was the case before the three experimental trials were recoded, these results do not indicate strong evidence for participants having a response bias to select either the streaky or steady option within each of the window sizes.

Pairwise comparisons of the response bias *c* distributions as a function of window size showed that the posterior probability that response bias for Window Size 1 is larger than Window Size 4 is 0.82. The posterior probability that response bias for Window Size 16 is larger than Window Size 4 is 0.76. And the posterior probability that response bias for Window Size 1 is larger than Window Size 16 is 0.64. These results do not indicate strong evidence for a difference in response bias *c* between each of the window sizes after recoding.

6.2.3. Individual Participants after Recoding

Individual differences in this discrimination task were also explored after the three experimental trials were recoded. After plotting each participant's *d'* and response bias value *c* for each window size after recoding, a very strong linear relationship was still found for Window Size 16 with the same correlation of .99 (see Figures 12). However, the correlation between *d'* and response bias decreased for Window Size 4 after recoding from .96 to .73 (see Figure 13). For Window Sizes 16 and 4, the top three performing participants (based on d') still showed a stronger tendency to see steady shooting (Bernoulli random). For Window Size 1 all participants showed a bias to select the steady option. The correlation between *d'* and bias increased after recoding from -.51

to -.69 (see Figure 14). This indicates that for Window Size 1, the weaker response bias a participant had the better they performed (based on *d'* values).

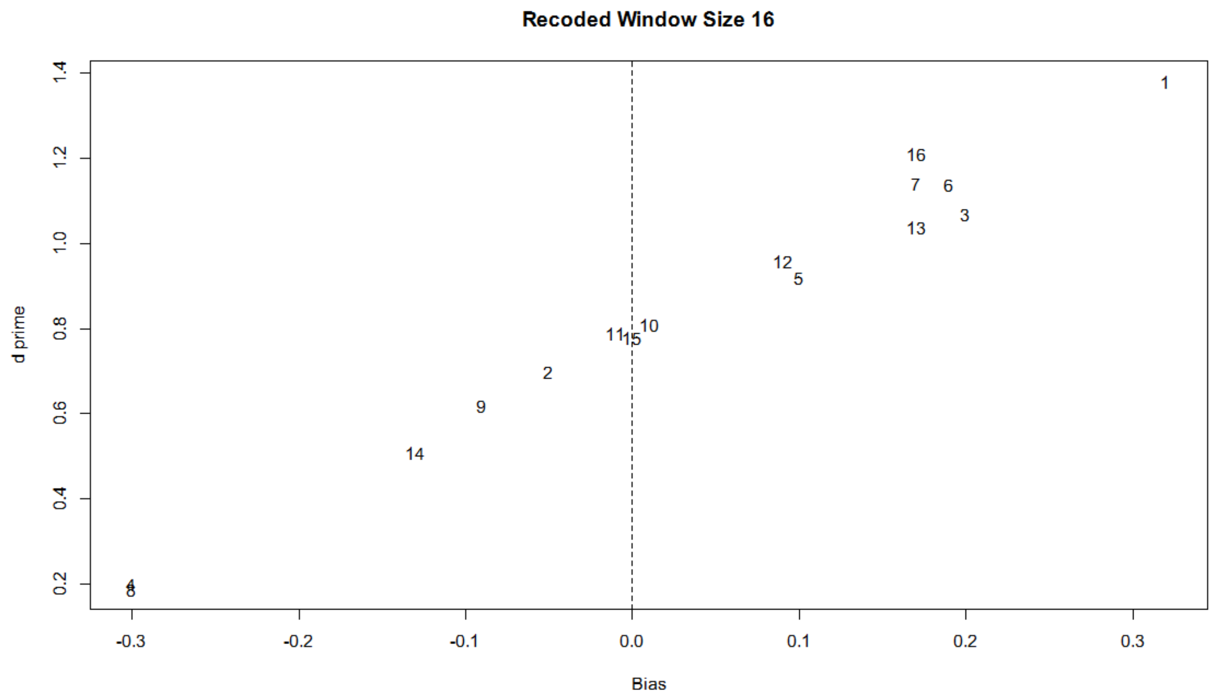


Figure 12. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 16 after the three experimental trials were recoded. Each plotted number indicates a participant.

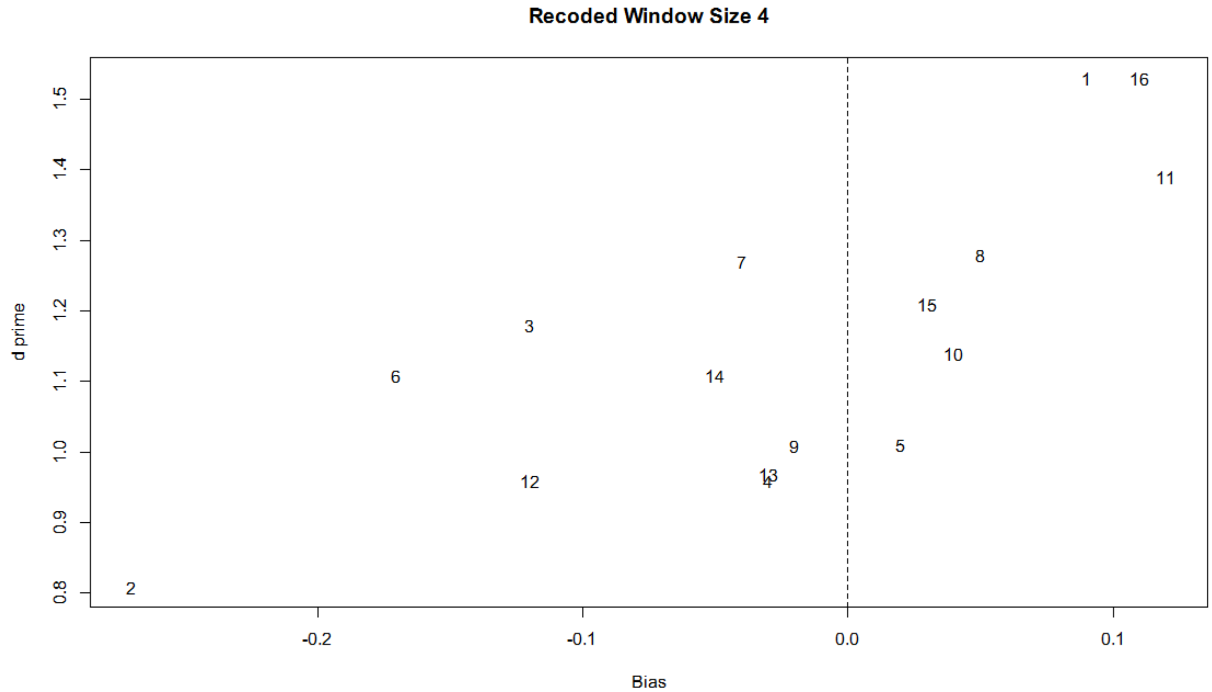


Figure 13. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 4 after the three experimental trials were recoded. Each plotted number indicates a participant.

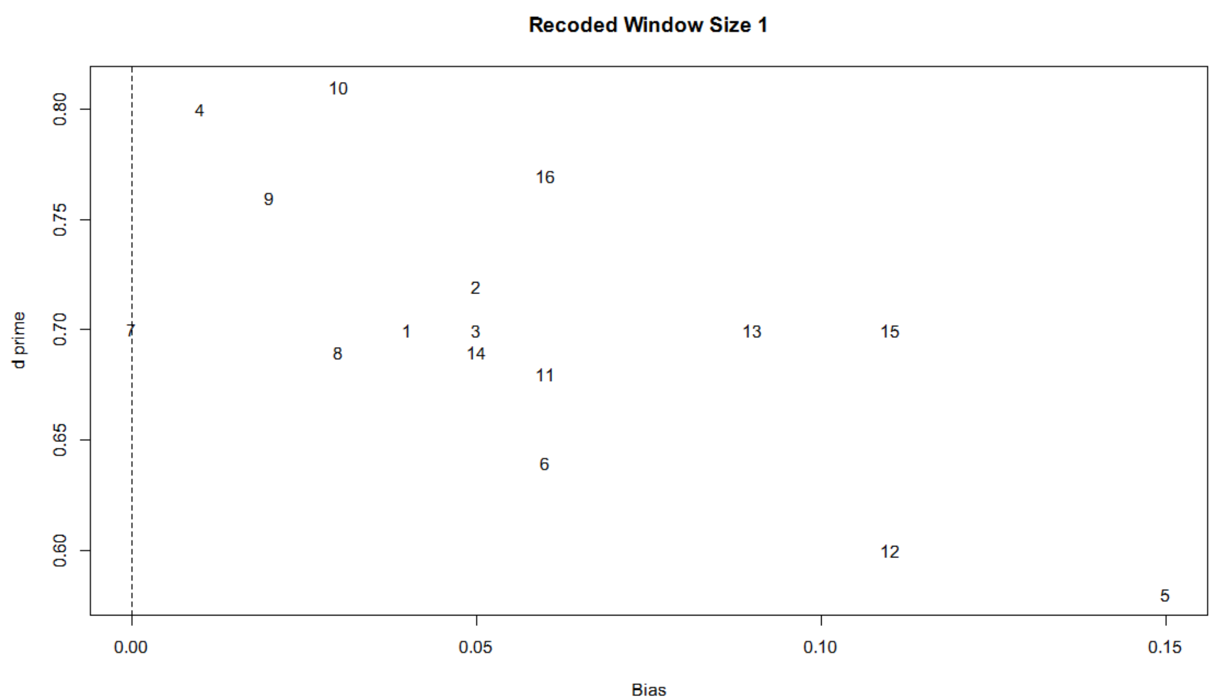


Figure 14. Graph of each participants' *d'* value plotted against their response bias *c* value for Window Size 1 after the three experimental trials were recoded. Each plotted number indicates a participant.

7. ARCHIVAL NBA DATA METHOD

Based on GVT's claim that a basketball player's shooting performance is analogous to the flips of a coin, the effect of home versus away game performance was explored in order to investigate potential contextual effects on NBA players' performance. One point guard from each of the 30 NBA teams was selected and their attempted and made field goals (excluding free throws) throughout the 2010-2011 regular season was recorded and the variance in the number of points scored for their home vs. away games was then analyzed using traditional logistic regression.

Selecting a point guard from each team began with eliminating the players that had played for more than one team during the 2010-2011 regular season. From the remaining point guards, the one player per team with the highest combined number of games played (*M* = 77) and average number of minutes played per game (*M* = 30.16) was chosen for the analysis. The following formula was used for the logistic regression, with an individual parameter for each player and a parameter indicating whether the shot accuracy was for a home or away game:

$$p = \frac{1}{1+e^{-(b_0+\sum_{j=1}^{k} b_j x_j)}}$$

8. ARCHIVAL NBA DATA RESULTS

The variance in the number of points scored did not significantly differ for players at home versus away games $\chi^2(1, N = 2309) = 1.57, p = .21$. The only significant predictor of shot accuracy was the player $\chi^2(29, N = 2309) = 44.97, p = .03$. These results did not provide support for the home-field advantage; however, in an extensive sports meta-analysis (discussed further in the Archival NBA Date segment of the Discussion section), Jamieson (2010) did find a significant advantage for home-field games.

9. DISCUSSION

9.1. Monte Carlo Simulations

Due to the criticisms for the lack of power in GVT's statistical tests, which often cite the power as inadequate for the conclusions drawn (Wardrop, 1999; Miyoshi, 2000; Korb and Stillwell, 2003), various amounts of data points were used for the Wald-Wolfowitz Runs Tests and spectral analyses for all of the Monte Carlo simulations. One sample size of particular interest is N = 512. This sample size is the closest sample size to what GVT used in their analysis of the NBA Philadelphia 76ers (*M* = 422.3) shooting performance while still being a power of two (allowing it to be analyzed by both the spectral analysis and the Runs Test).

A few of the simulation conditions had noteworthy results when the weakest sample size of N = 512 was used. In Quadrant 1, the p(hit) = 0.8 condition was the only condition in that quadrant to show deviations from the expectations of a Bernoulli process, which occurred only for the spectral analysis using N = 512. Furthermore, the 25% chance of alternation condition in Quadrant 2 and the simple learning curve condition with k = .001 in Quadrant 3 did not show deviations from the expectations of a Bernoulli process only when the weakest sample size of N = 512 was used for the Runs Test.

For the performance dependent learning curve simulations where the initial and asymptotic probabilities of success were .33 and .61 (respectively) and k = .005, the only condition to show deviations from the expectations of a Bernoulli process occurred with the weakest sample size of N = 512 when analyzed with the spectral analysis. And finally, for the performance dependent learning curve simulations where the initial and asymptotic probabilities of success were .20 and .80 and k = .005, the only condition to *not* show deviations from the expectations of a Bernoulli process was when the weakest sample size (N = 512) and the end 800 data points were analyzed using the Runs Test analysis.

It was anticipated and found that the weaker sample sizes produced inconsistent results when compared with the stronger sample sizes used. Furthermore, the higher learning rate conditions analyzing the end 800 events in the sequence seemed to be less likely to show deviations from the expectations of a Bernoulli random process, presumably because these sequences reach a plateau in performance (reach asymptotic probability of success) over the smallest span of trials, thus creating less variation in the probability of success at the end of the sequence.

9.2. Empirical Experiment

After the Bayesian model comparison resulted in a Bayes Factor of less than one, one experimental trial was recoded from being a steady sequence (Bernoulli random) to a streaky sequence (constrained) and two trials were recoded from being streaky sequences to being steady sequences. Although there is not strong evidence for a difference in *d'* before and after the three experimental trials were recoded (the posterior probability that the overall d prime across all window sizes after recoding is larger than before recoding is 0.64), to the extent that there is a difference, the trend suggests that model comparisons may be a better representation for human decision making based on the observed increase in *d'* values after these trials were recoded.

As a result of recoding the correct answer for the three experimental trials post hoc, there is a potential effect of giving participants incorrect feedback during the experiment on their answer selection. Although there were only 3 out of 90 experimental trials that could have given incorrect answer feedback, it would have been ideal to have replaced these three experimental trials before running the experiment in order to completely eliminate any potential influence incorrect feedback may have had on participant responses.

Based on the hypothesis that people have an evolutionary predisposition to see clumping or patterns of resources, potential participant response bias *c* was explored. Although it was hypothesized that participants would show a significant bias *c* to choose the streaky (constrained/structured) option rather than the steady (Bernoulli random) option, not finding a reliably different response bias *c* than zero when taken across participants was not completely unexpected. Because participants completed this task in a laboratory (artificial) setting, it is hard to rule out the possibility that participants could have altered their responses based on the environment and the form this task was given to them in.

Furthermore, it is important to note that for this experiment the number of streaky and steady trials was kept evenly distributed with 45 streaky trials/players and 45 steady trials/players (with the exception of the three recoded trials, which were recoded after all participants completed the experiment). Although the participants were not explicitly told the exact number of streaky and steady players they would be discriminating between, in an experimental setting not only could participants have anticipated that there were approximately an even ratio of streaky and steady players, but the immediate feedback they were given after each trial could have also potentially given them hints about the ratio of streaky to steady players. In a real world scenario, where the proportion of streaky and steady basketball players would be unknown, a valid hypothesis would still be that the spectators would show a bias towards seeing structure/constraints in player performance.

Another interesting finding with regard to response bias was that with Window Size 16 and 4 there was a very strong positive linear relationship between response bias and *d'* values. More specifically, the stronger a participant's bias to respond steady was, the higher their *d'* value. I would hypothesize that the plot of these variables would produce a negative parabola shape if there were participants with even stronger biases to respond steady. In other words, *d'* values and thus participant performance would very likely start to decrease once a particular response bias value was reached and surpasses.

However, even with relatively small response bias *c* values, there was still a very strong positive correlation between response bias and *d'* values for Window Sizes 16 and 4. This could indicate that the steady sequences were easier to determine than the streaky sequences since the participants who had a stronger bias to respond steady also had the highest sensitivity indices.. Again though, the group response bias *c* across window sizes was not significantly different from zero.

A final discussion topic with regards to the empirical experiment is the experiment completion time. The time needed to complete this experiment ranged from 50-140 minutes (average of 86 minutes). In contrast, the completion time for Cooper's (2013) experiment ranged from 20-45 minutes. The only substantial variation this experiment had from Cooper's (2013) was that this experiment contained post block questions (see Appendix C). However, participants typically took around 10-15 minutes to answer all of the questions presented in Appendix C. Even with the answering time for the post block questions taken out there is still quite a substantial difference in experiment completion time. Since both this experiment and Cooper's (2013) found that humans are able to accurately discriminate between Bernoulli random and alternatively constrained sequences, perhaps slight variation in experimenter instruction caused this fluctuation in experiment completion time.

9.3. Archival NBA Data

Although a valid place to explore possible contextual effects on basketball player's home vs. away game performance, the analysis of the NBA point guards' success rate for the 2010-2011 regular season did not provide statistically significant support for the home-field advantage. However, a meta-analysis by Jamieson (2010) found that athletic teams did in fact have a significant advantage for home-field games, with a moderator effect from time era (e.g., pre-1950, 1951-1970, 1971-1990, etc.), season length (less than 50 games, 50 to 100 games, more than 100 games), game type (regular season vs. championship), and sport (e.g., baseball, basketball, golf, etc.).

Furthermore, Willoughby (2014) explored the effect of the home-field advantage for teams in the MLB, NBA, NFL, and NHL. He found an impact on team performance and support for the home-field advantage based on the attendance rate of a team's home games. More specifically, a nonlinear relationship was found, such that teams with higher attendance rates for home games performed better in front of larger crowds, and teams with lower attendance for home games performed better in front of smaller crowds. It appears that there are indeed quite a few contextual variables that can impact a player's performance. With a larger sample size and a wider range of variables, perhaps the home-field advantage would have been further supported in the archival NBA analyses discussed in this paper.

10. SUMMARY AND CONCLUSIONS

10.1. Specific Implications Relative to Gilovich, Vallone, and Tversky (1985)

The results of the Monte Carlo simulations analyzed with different sample sizes provided valuable insights into the effect sample size has on significance results when using the Runs Test and Fourier spectral analysis. Several of the conditions provided inconsistent results when using the weakest sample size (N = 512) in comparison to the larger sample sizes ($N \geq 800$). These findings provide specific conditions/constraints on performance where a sample size of 512, comparable to GVT's sample size of N = 422.3, may be insufficient to provide an accurate discrimination of Bernoulli random from alternatively constrained sequences when using the Wald-Wolfowitz Runs Test and/or Fourier spectral analysis.

10.2. Specific Implications Relative to the Hot Hand

Validation for the belief in the hot hand and the use of heuristics as an adaptive strategy was explored through a human discriminability task using known constrained sequences in a laboratory setting. Performance dependent learning constrained binary sequences using the initial p(hit) = .33, the asymptotic p(hit) = .61, and k = .005 were *not* considered to significantly deviate from the expectations of a Bernoulli random process when the beginning 800 data points were analyzed using the Wald-Wolfowitz Runs Test and Fourier spectral analysis. Interestingly however, when this same constraint was used to generate sequences for the empirical experiment, the results indicate that people *are* able to make the distinction between the Bernoulli random and alternatively constrained sequences. Ensuing these empirical results, it follows that perhaps people are also able to detect patterned/constrained processes in a real-world setting (e.g., streaks in basketball performance), thus supporting the belief in the hot hand.

10.2.1. Pattern Perception: Fallacy or Adaptive Strategy?

The belief in the hot hand seems like a smart, adaptive strategy. Due to our evolutionary predisposition to see clumping or patterns of resources (Wilke & Barrett, 2009), people's belief in the hot hand seems like a scenario where people are able to leverage their pattern recognition abilities. Because of the vast amount of information available in the real-world, there is ample opportunity for people to potentially pick up on patterns or cues that can lead them to more advantageous choices. For example, the continual decreased pace of a player may indicate fatigue which, if detected by the opposing team, could be a rewarding opportunity to capitalize on (it is easier to score points when going head to head with a fatigued player).

It is important to provide an example that illustrates the advantage of having a default assumption of structure or patterns existing in the real world, as opposed to a default assumption of an unconstrained, unstructured world. Take a scenario where the goal is to predict the roll of a six-sided dice. There are two main aspects to consider, the first is how you *believe* this system to behave and the second is how this system *actually* behaves (further illustrated in Figure 15). We will consider two primary possibilities for human biases, or human assumptions about how this system (rolling a dice) is thought to behave. The first option is that the dice is assumed to be a standard six-sided dice, where all outcomes are equally likely (i.e., the probability of rolling a 1 is the same probability as rolling a 2, 3, 4, 5, or 6) and the dice is thus believed to be an unconstrained process.

The alternative option is to believe that all outcomes are *not* equally likely and that there is some sort of constraint to the outcomes of the dice (e.g., this is a loaded dice where there is the highest probability of rolling a 2).

Moreover, there are two different possibilities for the true process by which the dice operates (i.e., how does it actually behave). The first option is that this dice is actually an unconstrained process, where all outcomes are equally likely and there is maximum uncertainty in the outcomes of rolling this dice. The second option is that this dice is actually a constrained process, where all outcomes are *not* equally likely and the opportunity exists for you to optimize, or more accurately predict the outcomes of rolling the dice.

In order to make the argument that it is more advantageous to have a default assumption/bias that natural systems in the world behave in a constrained (patterned) way, as opposed to an unconstrained (random) way, we must consider what there is to gain and lose with each belief (illustrated in Figure 15). Let us begin with a scenario where the human believes the dice to be just a regular, unconstrained six-sided dice, where the belief is that there is no possibility of better predicting the outcome of any given roll of the dice since all outcomes are equally likely. In this scenario however, let us say that this assumption is wrong and that the dice actually has some sort of constraint on its performance (e.g., it is a loaded dice where the probability of rolling a 2 is the highest). With this scenario, the human bias would have eliminated any potential predictability advantage, since there is no point in trying to predict/leverage events of a system that is believed to be unconstrained. With the current human assumption of the system being unconstrained, it would result in a missed opportunity to make a better outcome prediction.

Maintaining the human's assumption that the dice is unconstrained, say this assumption is indeed correct and the dice is in fact an unconstrained, unpredictable process. In this scenario, although the human assumption is correct, there is nothing to be gained from having an unconstrained bias/assumption of the system, since there is no predictive advantage to possibly be gained from a system that behaves unpredictably (i.e., unconstrained).

Now switch the human bias so that their assumption is that the six-sided dice behaves in some sort of constrained/patterned way. If the dice is actually an unconstrained process in this scenario, having the assumption that it behaves in a constrained way would not have any negative effect because again, you cannot better predict events/outcomes of a system that is unconstrained. However, if this dice is actually constrained in some way and your assumption is that it behaves as such, there is an opportunity for you to leverage the constraints of the system (i.e., make more accurate predictions about the outcomes of dice rolls).

After playing out each of these scenarios, it seems abundantly clear that the more advantageous default bias/assumption about the world is that there is some sort of pattern to be leveraged. If you are wrong, there is nothing lost since you cannot predict events of an unconstrained system anyway, but if you are right, there is a potential that you can leverage the constraints of the system.

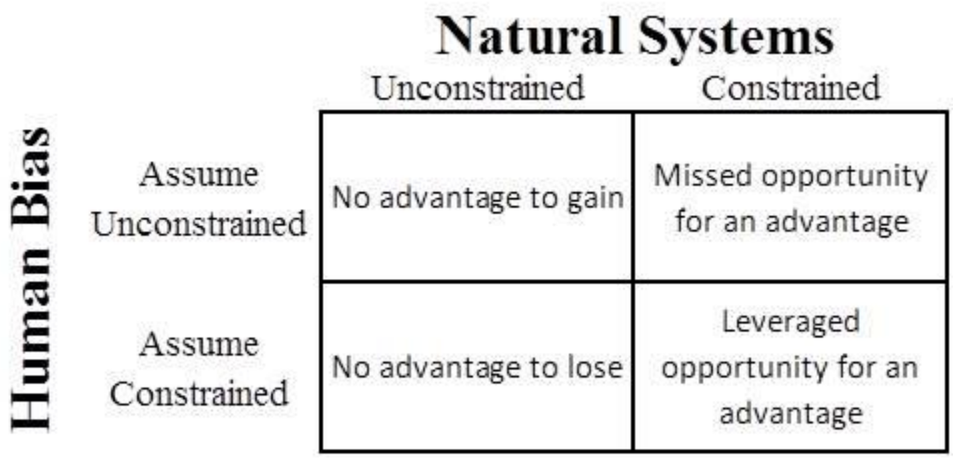


Figure 15. A matrix showing the possible outcomes from human assumptions about how natural systems behave combined with the actual way these natural systems behave.

10.3. General Implications Relative to Human Decision Making

10.3.1. Heuristics: Bias or Smart Mechanism?

Is belief in the hot hand a bias, an error, or an effective strategy for detecting patterns in nature? Conventionally heuristics tend to be analyzed in artificial situations specifically structured around normative logic that result in numerous scenarios where heuristics are shown to be erroneous (Lopes & Oden, 1991). However, when heuristics, such as the ones proposed by Tversky and Kahneman (1974), are employed in real-world scenarios, they may actually be properties of abductive logic that allow people to make ecologically valid guesses in unknown situations. It appears that in real-world situations the belief in the hot hand, and more broadly the assumption that the world is in fact made up of many constrained processes (as opposed to many random ones), seems like a smart default "bias."

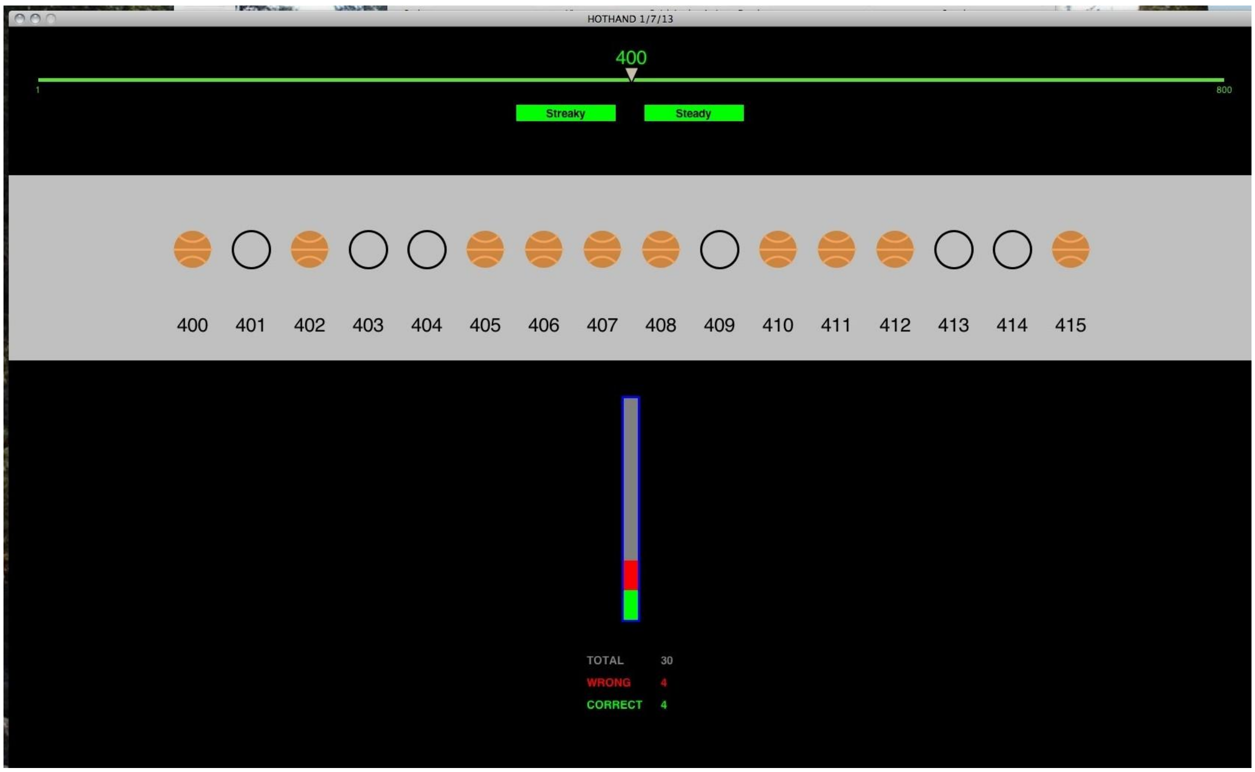
10.3.2. Abduction and Ecological Rationality

Cooper (2013) proposed that "…there is an alternative perspective on human reasoning that has roots in early functional/pragmatic approaches to human cognition. For example, Peirce (1877/1997) offered the construct of abduction as an alternative to classical logic. Abduction is an approach to rationality that is grounded in the practical success of beliefs, rather than in the syntax of arguments" (p. 39). In abduction systems, the validity of a belief is judged relative to whether the belief has led to satisfactory outcomes in the past. In other words, beliefs are reinforced or validated by success. Thus, for example, beliefs in the hot hand might be tested relative to the success of the coaches (like Phil Jackson) who adhere to this belief. An implication of this is that rather than looking to normative mathematics as the measure for rationality, look to the practices (and beliefs) of the experts (people who are successful).

More recently, the term ecological rationality is used to describe the value of our beliefs and cognitive processes as being neither good nor bad per se, but ecologically valid dependent on the goals the decider seeks to attain in that context (Todd & Gigerenzer, 2012; Gula & Raab, 2004). Cooper (2013) suggested that "From the perspective of Ecological Rationality, heuristics are considered to be analogous to Runeson's construct of *smart instrument*. That is, the use of heuristics reflects an attunement to structure (invariants) in natural ecologies. Thus, it may be an example of an abductive form of rationality that leverages constraints in the problem ecology in ways that support successful adaptations" (p. 39). This research explored the idea that the belief in the hot hand is an effective, adaptive strategy rather than an illusion or neglect of probability theory.

APPENDIX A

Empirical Experiment Interface



APPENDIX B

Participant Instructions

Imagine that you are a scout for a reputable college basketball team. You have been asked by the coach to identify high school players that are going to fit the needs of the program. You will be reviewing shooting performance of players in their final year of high school, and will be responsible for deciding whether they fit the needs of the program. You will be reviewing the sequence of 800 shots that were taken over the season. You will then have to decide what type of player's sequences you are viewing and decide amongst 2 options that are presented to you. In the sequences that you are presented, you will have a viewing window where you can view 16, 4, or 1 event at a time. Once you have reviewed the sequence, you will be presented with two choices. You must make the appropriate selection based on the nature of the sequence that you have just examined.

You will be examining players who exhibit learning potential but are also streaky. These players steadily improve across the season, but are somewhat inconsistent in their performance from game to game. We would expect to see streaky behavior with shots (hits and misses), but those streaks should typically improve as the season progresses. We would be interested in looking at this type of player for our tem. Please correctly identify the streaky shooters who improve as the season progresses from the players with a consistent shooting percentage (0.440).

APPENDIX C

Participant Debriefing

End of Block 1

1. What was your strategy for discriminating the sequences?
2. Why did you use this strategy?
3. Rate your level of engagement (how interesting did you find the task):

(Not at all engaged) 1 2 3 4 5 6 7 (Very engaged)

4. Rate your level of motivation to correctly identify the sequences:

(Not at all motivated) 1 2 3 4 5 6 7 (Very motivated)

5. Do you think you could identify players more accurately if given more trials?
6. Was it easier to identify the steady shooters or the streaky shooters?

End of Block 2

1. What was your strategy for discriminating the sequences?
2. Why did you use this strategy?
3. Rate your level of engagement (how interesting did you find the task):

(Not at all engaged) 1 2 3 4 5 6 7 (Very engaged)

4. Rate your level of motivation to correctly identify the sequences:

(Not at all motivated) 1 2 3 4 5 6 7 (Very motivated)

5. Do you think you could identify players more accurately if given more trials?
6. Was it easier to identify the steady shooters or the streaky shooters?

End of Block 3

1. What was your strategy for discriminating the sequences?
2. Why did you use this strategy?

3. Rate your level of engagement (how interesting did you find the task):

(Not at all engaged) 1 2 3 4 5 6 7 (Very engaged)

4. Rate your level of motivation to correctly identify the sequences:

(Not at all motivated) 1 2 3 4 5 6 7 (Very motivated)

5. Do you think you could identify players more accurately if given more trials?
6. Was it easier to identify the steady shooters or the streaky shooters?
7. Which window size did you find it the ***easiest*** to discriminate between sequences?
8. Which window size did you find it the ***hardest*** to discriminate between sequences?
9. Any suggestions to make the discrimination task easier?
10. Other feedback:

REFERENCES

Alter, A. L., & Oppenheimer, D. M. (2006). From a fixation on sports to an exploration of mechanism: The past, present, and future of hot hand research. *Thinking & Reasoning, 12(4)*, 431-444.

Avugos, S., Köppen, J., Czienskowski, U., Raab, M., & Bar-Eli, M. (2013). The "hot hand" reconsidered: A meta-analytic approach. *Psychology of Sport and Exercise 14*, 21-27.

Bocskocsky, A., Ezekowitz, J., & Stein, C. (2014). The hot hand: A new approach to an old "fallacy". *MIT Sloan Sports Analytics Conference*, 1-10.

Compagner, A. (1991). Definitions of randomness. *American Journal of Physics, 59*, 700-705.

Cooper, J. (2013). Heuristics: Bias vs. smart Instrument. An exploration of the hot hand. *Unpublished doctoral dissertation*, 1-114.

Falk, R. (1991). Randomness—An ill-defined but much needed concept. Commentary on "Psychological conceptions of randomness.". *Journal of Behavioral Decision Making, 4*, 215-218.

Ford, J. (1983, April). How random is a coin toss? *Physics Today*, 40-47.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *The Annual Review of Psychology, 17*, 451-482.

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

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

Gula, B., & Raab, M. (2004). Hot hand belief and hot hand behavior: A comment on Koehler and Conley. *Journal of Sport & Exercise Psychology, 26*, 167-170.

J, A. A. (1965). Chance. *Scientific American, 213(4),* 44-54.

Jamieson, J. P. (2010). The home field advantage in athletics: A meta-analysis. *Journal of Applied Social Psychology*, 1819-1848.

Kac, M. (1983). Marginalia: What is random? . *American Scientist, 71*, 405-406.

Kolata, G. (1986). What does it mean to be random? *Science, 231*, 1068–1070.

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

Lopes, L. L., & Oden, G. C. (1991). The rationality of intelligence. *Poznark Studies in the Philosophy of the Sciences and the Humanities, Vol.21*, 199-223.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. *Cambridge: Cambridge University Press*.

Miller, J. B., & Sanjurjo, A. (2014). A cold shower for the hot hand fallacy. *Unpublished manuscript*, 1-57.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review 109(2)*, 330-357.

Raab, M., Gigerenzer, G., & Gula, B. (2012). The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied 18(1)*, 81-94.

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.

Shah, A., & Oppenheimer, D. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin, Vol. 134, No. 2,*, 207–222.

Smith, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin & Review 10(3)*, 753-758.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185(4157),* 1124-1131.

Wilke, A., & Barrett, C. H. (2009). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior 30*, 161-169.

Willoughby, J. (2014). Attendance, home advantage, and the effect of a city on its professional sports teams. 1-23.

Yaari, G., & David, G. (2012). ‘‘Hot Hand’’ on strike: bowling data indicates correlation to recent past results, not causality. *PLoS ONE 7(1): e30112.doi:10.1371/journal.pone.0030112*.