2013

Heuristics: Bias vs. Smart Instrument. An Exploration of the Hot Hand

Jehangir Cooper

Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the Industrial and Organizational Psychology Commons

Repository Citation
https://corescholar.libraries.wright.edu/etd_all/751

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.
HEURISTICS: BIAS VS. SMART INSTRUMENT. AN EXPLORATION OF THE HOT HAND

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

By

JEHANGIR COOPER
M.S., Wright State University, 2010

2013
Wright State University

John Flach, Ph.D.
Dissertation Director

Scott Watamaniuk, Ph.D.
Graduate Program Director

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

R. William Ayers, Ph.D.
Interim Dean, Graduate School

Committee on Final Examination

John M. Flach, Ph.D.

Kevin B. Bennett, Ph.D.

Valerie L. Shalin, Ph.D.

Clark A. Shingledecker, Ph.D.
ABSTRACT


Classical perspectives on judgment and rationality view heuristics as erroneous, leading to suboptimal judgments. Conversely, ecological perspectives view heuristics as smart mechanisms that result in good judgments in the face of uncertainty. Our research focused on the hot hand heuristic and examined it using non-linear analysis methods. This research attempted to answer two questions. The first question concerned the applicability of frequency analysis methods for detecting constraints (such as the hot hand) or structure in a time series of binary data, which we attempted to investigate through Monte Carlo simulations. We found that this method was sensitive enough to detect structure. The second question was concerned with whether humans are able to discriminate random series from constrained (structured) series. We conducted an experiment which investigated whether time series validated by the frequency analysis as constrained were detectable by humans. Our results showed that humans have an ability to recognize a constrained series more often than chance. A link between the strength of constraints in the spectral analysis with performance in discrimination of the task was demonstrated, suggesting that the higher the strength of constraint (according
to the spectral plots) the easier it is to discriminate. The spectral plots might hold validity to index the psychophysics of pattern detection in humans. Our results gives credence to the ecological justification for the use of heuristics (such as the belief in the hot hand), especially in a skilled situation, such as sports performance.
# TABLE OF CONTENTS

1. INTRODUCTION .................................................................................................................. 1  
   1.1. Heuristics: Bias versus Smart Instrument ............................................................... 1  
   1.2. The Gambler’s Fallacy and Hot Hand Phenomenon ................................................. 3  
   1.3. Is performance Streaky? .......................................................................................... 6  
   1.4. The Research Approach .......................................................................................... 11  

2. MONTE CARLO SIMULATIONS ...................................................................................... 19  
   2.1. Frequency Analysis .................................................................................................. 19  
   2.2. Method .................................................................................................................... 22  
      2.2.1. Design and Procedure ..................................................................................... 22  
      2.2.2. General Expectations ...................................................................................... 26  
   2.3. Results .................................................................................................................... 27  
      2.3.1. Outline .............................................................................................................. 27  
      2.3.2. Independent/Stationary Processes (Quadrant 1) .............................................. 28  
      2.3.3. Independent/Nonstationary Processes (Quadrant 2) ....................................... 33  
      2.3.4. Dependent/Stationary Processes (Quadrant 3) ................................................. 38  
      2.3.4.1. Learning ....................................................................................................... 38  
      2.3.4.2. Local Dependence ....................................................................................... 41  
      2.3.5. Dependent/Nonstationary Processes (Quadrant 4) ........................................... 46  
      2.3.6. General Conclusions ....................................................................................... 49  
      2.3.7. Link Between Beta Slopes and Absolute Z Values ......................................... 49  
   2.4. Discussion ................................................................................................................. 50  
      2.4.1. Quadrant 1 ........................................................................................................ 51  
      2.4.2. Quadrant 2 ........................................................................................................ 52  
      2.4.3. Quadrant 3 ........................................................................................................ 54  
      2.4.4. Quadrant 4 ........................................................................................................ 56
3. SERIES DISCRIMINATION TASK .................................................................................................................. 60
   3.1. Introduction .................................................................................................................................................. 60
      3.1.1. Conceptions of Randomness .................................................................................................................. 61
      3.1.2. Previous Research on Human Judgment .................................................................................................. 62
      3.1.3. Our Research Approach ....................................................................................................................... 64
   3.2. Method .......................................................................................................................................................... 65
      3.2.1. Design .................................................................................................................................................. 65
         3.2.1.1. Independent/Stationary Processes (Quadrant 1) ........................................................................... 66
         3.2.1.2. Independent/Nonstationary Processes (Quadrant 2) ................................................................. 66
         3.2.1.3. Dependent/Stationary Processes (Quadrant 3) ........................................................................... 67
         3.2.1.4. Dependent/Nonstationary Processes (Quadrant 4) ................................................................. 68
      3.2.2. Participants ......................................................................................................................................... 69
      3.2.3. Equipment ......................................................................................................................................... 69
      3.2.4. Procedure .......................................................................................................................................... 70
      3.2.5. Analysis ............................................................................................................................................ 70
      3.2.6. General Expectations .......................................................................................................................... 74
   3.3. Results ....................................................................................................................................................... 75
      3.3.1. Outline .............................................................................................................................................. 75
      3.3.2. Independent/Stationary Processes (Quadrant 1) ................................................................................ 75
      3.3.3. Independent/Nonstationary Processes (Quadrant 2) ........................................................................ 80
      3.3.4. Dependent/Stationary Processes (Quadrant 3) ................................................................................ 80
         3.3.4.1. Learning ...................................................................................................................................... 82
         3.3.4.2. Local Dependence ...................................................................................................................... 86
      3.3.5. Dependent/Nonstationary Processes (Quadrant 4) ........................................................................... 89
      3.3.6. Summary of Average Trial Time and Scanning Behavior ................................................................. 92
   3.4. Discussion ................................................................................................................................................... 95
4. GENERAL DISCUSSION................................................................................................................................. 97

4.1. d’ versus Beta Slopes and Runs Tests ................................................................................................. 97

4.2. Implications................................................................................................................................................. 100

4.3. Future Directions of the Research ........................................................................................................ 104

APPENDIX A.................................................................................................................................................. 108

APPENDIX B .................................................................................................................................................... 113

REFERENCES................................................................................................................................................... 117
LIST OF FIGURES

Figure 1. Nature of generating processes of events that can occur, divided into four Quadrants ............................................................................................................................................................................................... 13

Figure 2. Spectral plot of white noise condition ........................................................................................................................................................................................................ 20

Figure 3. Spectral plot of pink noise condition ..................................................................................................................................................................................................... 21

Figure 4. Process for generating Beta slopes from raw data ......................................................................................................................................................... 25

Figure 5. Moving average for sample simulation in Quadrant 1 using a window of 20 samples to compute the average ................................................................................................................................................................................................ 29

Figure 6 Spectral plot of a sample simulation for Quadrant 1 ........................................................................................................................................................................ 31

Figure 7a. Moving average for sample simulation in Quadrant 2 with 10% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples ........................................................................................................................... 34

Figure 7b. Moving average for sample simulation in Quadrant 2 with 25% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples ........................................................................................................................... 34

Figure 7c. Moving average for sample simulation in Quadrant 2 with 50% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples ........................................................................................................................... 34

Figure 8. Spectral plot of a sample simulation for Quadrant 2 (10% chance of alternation) .................................................................................................................................................................................................................. 36

Figure 9. Moving averages for sample simulations of learning curves in Quadrant 3 using a window of 20 samples ................................................................................................................................................................................................ 39
Figure 10. Spectral plot of a sample simulation for learning curves in Quadrant 3
   \( k=0.001 \) ........................................................................................................... 40

Figure 11. Moving averages for sample simulations of shot dependency in Quadrant 3
   using a window of 20 samples .................................................................................. 42

Figure 12. Spectral plot of a sample simulation for one shot dependency in Quadrant 3
........................................................................................................................................... 44

Figure 13. Spectral plot of a sample simulation for a five shot dependency in Quadrant 3
........................................................................................................................................... 45

Figure 14. Moving average for a sample simulation in Quadrant 4 using a window of 20 samples
........................................................................................................................................... 47

Figure 15. Spectral plot of a sample simulation in Quadrant 4, \( k=0.002 (+/-0.1) \) ..... 48

Figure 16. Relationship between Beta slopes (frequency analysis) and absolute Z value (runs test)
........................................................................................................................................... 50

Figure 17. Screenshot of the experimental condition program............................................ 70

Figure 18. Group d' obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for Quadrant 1 ................................................................. 77

Figure 19. Group d' obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for Quadrant 2 ................................................................. 82
Figure 20. Group d’ obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for the learning curve condition in Quadrant 3 ................................................................. 85

Figure 21. Group d’ obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for the shot dependency condition in Quadrant 3 ................................................................. 88

Figure 22. Group d’ obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for Quadrant 4 ................................................................. 91

Figure 23. Relationship between frequency analysis (Beta slopes) and experimental performance (d’). .......................................................................................................................... 98

Figure 24. Relationship between the runs test (|Z| value) and the experimental performance (d’)............................................................................................................................. 99
LIST OF TABLES

Table 1  Results of the Simulations Conducted for all Four Quadrants .......................... 30
Table 2  Series Discrimination Performance (% correct and d’ for each window size)
in Quadrant 1 .................................................................................................................. 76
Table 3  Paired Comparisons Between the Window Sizes in Each Quadrant .............. 78
Table 4  Series Discrimination Performance (% correct and d’ for each window size)
in Quadrant 2 .................................................................................................................. 80
Table 5  Series Discrimination Performance (% correct and d’ for each window size)
for the simple learning curve condition in Quadrant 3 .............................................. 83
Table 6  Series Discrimination Performance (% correct and d’ for each window size)
for the 5 shot dependency condition in Quadrant 3 .................................................... 87
Table 7  Series Discrimination Performance (% correct and d’ for each window size)
in Quadrant 4 .................................................................................................................. 90
Table 8  Average Trial Time and Scanning Variance for Each Window Size for Each
Quadrant .......................................................................................................................... 93
1. INTRODUCTION

1.1 Heuristics: Bias versus Smart Instrument

This research attempts to illuminate the differences between two different perspectives of human rationality – the rationalist tradition (or classical view of rationality) and the ecological alternative. We will be discussing the differing attitudes towards heuristics in decision-making between these two approaches. Heuristics are simple shortcuts to problem solving, and are used when a complete search for solutions is not viable. They are often used in situations of high complexity, or limited information.

The first perspective consists of ideas of rationality rooted in economic and statistical (logical) models. This is currently the dominant model in the scientific community; where normative prescriptions of rationality are considered to be standards for rational thinking. The works of Tversky and Kahneman (1974) have embodied the classical rationalist tradition in their research on human decision-making and rationality. Tversky and Kahneman (1974) have described the use of heuristics in their work, and stated that while “heuristics are highly economical and usually effective,” using them can lead to “systematic and predictable errors” (p. 1131). Tversky and Kahneman (1974) also state that humans are prone to making
errors in judging probability as the heuristics employed might not reflect or account for the dynamics in the situation in which a solution is sought.

The second (ecological) perspective consists of the ideas of rationality rooted in the theory of abduction, which assesses the quality of decision-making logic with respect to the pragmatic value of the outcome (Frankfurt, 1958). In the ecological tradition, rationality involves the adaptive leveraging of the structure that is present in the task environment to satisfy the demands of the decision maker in the face of uncertainty. In other words, it involves the deliberate use of structure or patterns associated with specific contexts in order to make the best decision (based on the pragmatic value of the outcome) in an uncertain task environment. Therefore, in contrast to the rationalist tradition, the ecological perspective views heuristics as smart mechanisms (Runeson, 1977) that adapt to the situated context of the task environment in order to simplify the decision making process (and reduce the cognitive load when making decisions by bypassing traditional information processing channels). Gigerenzer (2008) has argued for a more ecological perspective on rationality. He states that heuristics “exploit our evolved capacities” and “can provide solutions to problems that are different from strategies of logic and probability” (p. 27). These intuitions are consistent with the idea that rationality emerges from a self-organizing dynamic involving a closed-loop coupling between agent and environment, as typically reflected in constructs such as ‘situated cognition’ (Suchman, 1987).
The rationalist tradition (Kahneman and Tversky) seeks to heighten awareness of the errors associated with so that people will follow the normative rules (based in statistics and economics) more consistently. The ecological tradition (Lindblom, Gigerenzer and Todd) argues that natural situations do not satisfy the assumptions of normative models. Lindblom (1959) talks about “muddling through” as a mechanism for arriving at solutions as an iterative process. In this process, feedback from the environment is received constantly as decisions are tested, until a satisfactory outcome is reached. Thus, smart decision making involves refining the muddling process to fully utilize the situated constraints.

To illuminate the difference between these 2 perspectives, this research focuses on a phenomenon that can be approached by both of these traditions – the hot hand. Unlike most of the empirical work on heuristics motivated by the classical rationalist tradition, the hot hand involves judgments based in a natural situation (usually sports performance – such as basketball).

1.2 The Gambler’s Fallacy and Hot Hand Phenomenon

Suppose you were flipping a coin and trying to guess the outcome of the flip – it’s either heads, or tails. Imagine that the coin is fair, which means that the probability of obtaining heads is the same as the probability of obtaining tails. Now suppose that this coin is flipped 5 times, and the outcome of each of those events is heads. On the next flip, many people might assume that the outcome would more
likely be tails than heads, since heads has already come up so many times, and since
the probability of obtaining tails is the same as obtaining heads (i.e., tails are “due
up”). This is called the gambler’s fallacy – where people mistakenly assume that
there is some sort of memory in a series of independent events (that is, people
assume that the previous results obtained have some sort of bearing on the next
result). Since we know that coin-flipping is independent, we can attribute the
gambler’s fallacy beliefs to an error or bias, which arises out of the
representativeness heuristic. The representativeness heuristic is a belief that a small
sample of events (flipping a coin a few times) should reflect the properties of the
larger distribution from which it was obtained (that flipping a coin 10 times should
yield an approximately even number of heads and tails, since it is a fair coin and has
equal probabilities of obtaining heads or tails).

While the gambler’s fallacy is an example of the belief in negative recency
(belief that the next event will be the opposite of the previous events), the hot hand
is an example of the belief in positive recency (belief that the next event will be the
same as the previous events). Like the gambler’s fallacy, the hot hand has been
explained in terms of the representativeness heuristic. Gilovich, Vallone and
Tversky (1985) explained that judgment of a sequence by the sample of
representativeness can also lead to a rejection that the sequence is in fact random.

The hot hand is the belief that performance in sports is not governed by a
random process based on simple probability alone – that is people have ‘hot’ and
‘cold’ streaks where the success of the next event is influenced by recent events. For
example, this suggests that a foul shooter in basketball will be more or less likely to make the current attempt depending on whether s/he has made or missed the last few shots. Many people with even a cursory interest in sports have suggested feeling at one point or the other, in their experience either watching a sports game or competing in one, that one of the players has gone on a “hot streak”. They believe that the player is performing better than expected, given their career averages. For example, while watching a basketball game, if we see a 0.300 career shooter hit the target on the last 9 attempts, we might consider the player to be on a hot streak. We would then expect that the player would be more likely to make the next shot, so it would be in the best interest of the team to allow that player on the hot streak to take that shot. Conversely, if the same player (0.300 career shooter) has missed the last 10 shots, many people would think that they would be more likely to miss the next shot, since they are in a “slump” (performing perceivably worse than would be expected).

In summary, believers in the gambler’s fallacy and the hot hand are making a judgment about the sequence of events, based on a small sample. In the gamblers fallacy, they expect that the number of heads and tails will ‘even out’ in the small sample. In the hot hand phenomenon, they decide that the process is constrained (i.e., not governed by random probability alone) and they invent constraints (i.e., hot or cold streak) to explain the properties of the small sample. Thus, these phenomena suggest that people seem to assume that properties of the small samples are representative of the larger sequence.
1.3 Is Performance streaky?

The first question with regard to the hot hand phenomenon is whether the actual shooting behavior is independent (i.e., operating under the same principles as flipping a fair coin) or whether there actually are constraints in performance as assumed by the hot hand belief. Constraints governing performance can manifest themselves as streaky behavior in time series.

There has been much research conducted to determine whether performance in sports is truly streaky (that is, do streaks appear more often than would be expected by random chance alone). The outcomes of the research have been mixed, and some of it will be explored below.

One of the most important works in the area of streaky performance in sports from the rationalist tradition is the work of Gilovich, Vallone and Tversky (1985), entitled “The Hot Hand in Basketball: On the Misperception of Random Sequences”. In this investigation Gilovich et al. analyzed the shooting records of the Philadelphia 76ers (a professional basketball team in the NBA) using serial probabilities and found that the likelihood of a successful shot did not necessarily increase following a previous successful shot. They also conducted an analysis on the Boston Celtics (another professional basketball team in the NBA), which yielded the same conclusion. Gilovich et al. (1985) also conducted a controlled shooting experiment with college students, and found that there was no positive correlation
between the outcomes of successive shots (that is, their performance was not streaky). However, interestingly, even though they found that performance was not streaky, most people had a perception that performance was streaky. Gilovich et al. attributed this to the law of small numbers, where people assume that a random sample drawn from a larger population would be representative of that population as a whole.

Although basketball shooting potentially involves many factors that would not be involved in a simple process like coin tossing (e.g., skill, confidence, fatigue), the results obtained by Gilovich et al. (1985) suggest that for skilled players, each shot is statistically independent from the prior shots. Thus, the belief in streaks appears to be a bias or illusion, based on the statistical conclusions they obtained. Tversky and Gilovich (2005) conclude that the belief in the hot hand (streaky behavior) could be a hindrance, as decisions to pass the ball to the “apparently” hot player are not justified. The inference is that the player who is most likely to make the next shot is the player with the highest overall shooting percentage (i.e., the career average is the best predictor of future performance). Koehler and Conley (2003) also discussed the notion of streaks in basketball. They analyzed the NBA Long Distance Shootout contest, and found that there was no substantial evidence of streaky performance, over and above what would be expected by chance. Like Tversky and Gilovich (2005), Koehler and Conley (2003) stated that the belief in streaky performance in sports could have costly implications. They recommend that coaches, managers and athletes should not use the ‘illusion of the hot hand’ to predict future performance, and instead predict it based on the athletes’ base rate of
success in similar circumstances. Attali (2013) reported that while the hot hand is illusory, it influences coaching decisions that decrease the likelihood of the player being substituted, whereas the player is more likely to take riskier shots and result in a decreased probability that the next shot is successful.

Albright (1993) conducted an analysis of the hitting streaks that occurred in baseball. Albright, while acknowledging that hitting streaks do occur in baseball, investigated whether these streaks occurred more frequently than they would under a probabilistic model of randomness. Albright found that while some players did exhibit some streakiness (as was expected) the behavior of the players, on aggregate, did not exhibit any unusually streaky behavior.

There have been several critiques against the normative notion of the hot hand in sports, particularly the findings of Gilovich et al. (1985). Sun (2004) pointed out that the analysis method was too primitive. Korb and Stillwell (2003) stated that a power analysis revealed their tests were weak, thus cautioning against the conclusions of Gilovich et al. (1985). Adams (1992) indicated the absence of true temporal data for examining streaks, and postulated that the temporal space between events could influence the perception of a streak (a successful shot would be more likely if less time has passed since a previous successful shot). Larkey, Smith and Kadane (1989) and Hooke (1989) stated that the complexity of different offenses and defenses were unaccounted for.

Gilden and Wilson (1995) explored the concept of streaky behavior in skilled performance. They found that performance in golf putting and darts tended to be
streaky. This violates the general assumption proposed by Gilovich et al. (1985) that streaky performance is just an illusion. Gilden and Wilson (1995) hypothesized that since the participants in this investigation were not experts (they were students recruited through advertising at their university) that there was a learning curve involved, and “a secular increase of hit rate over a block of trials could result in a deficit of runs relative to the expectation of a stationary Bernoulli process operating at the average hit rate” (p. 264). Also, consistent with common intuition, a delay between trials tended to make the outcomes of the trials independent, according to Gilden and Wilson (1995). Bower (2011) concedes that there might be some compelling evidence for streaky behavior in sports, as evidenced by Orel Hershiser (pitched 59 scoreless innings in a row in the 1988 MLB season), Wayne Gretzky (51 games with a point in the 1983-84 NHL season) and Johnny Unitas (47 consecutive games with a touchdown pass in professional football from 1956-1960). Adams (1995) stated that pocket billiards players experienced momentum (an enhanced psychological state and functioning) which resulted in a positive correlation in their performance. Smith (2003 and 2004) found that horseshoe pitchers and 10-pin bowlers also exhibited streaky behavior.

Bar-Eli, Avugos and Raab (2006) conducted a review of the research on streaky performance in sports. They found that the research was generally mixed – they identified 13 studies indicating non-streaky performance (performance consistent with a random Bernoulli process) and 11 studies indicating streaky performance.
However, there is a vein of research that validates the belief in the hot hand as an evolutionary adaptation to clumped resources. Wilke and Barrett (2009) suggest that the hot hand is a cognitive adaptation to help people predict the presence of items in space and time. For the purposes of foraging (which was an evolutionary necessity), the prediction that resources tend to be clumped in space and time was seen as an evolutionary advantage. Wilkes and Barrett argue that humans are evolutionally predisposed to seeing patterns as it provided them with an advantage in foraging in ancestral times. Scheibehenne, Wilke and Todd (2011) also argue that the belief in the hot hand is ecologically adaptive. They argue (along with Bower, 2011), that the theory of clumped resources, while originally applied to evolutionary behavior, such as foraging, is also prevalent in other domains, such as sports performance. Wilke and Barrett (2009) showed that a preference towards positive recency or clumps transcended culture and educational level, where their participants (32 students at UCLA and 32 Shuar hunter-horticulturalists, both distinct demographically) exhibited a bias towards streaky behavior in random coin flips. Asparouhova, Hertzel and Lemmon (2009) argued that people are more likely to believe in the hot hand as their perception of the randomness of the generating process decreases. Ayton and Fischer (2004) validate this notion in their findings that suggest when it comes to human performance, people tend to believe that a hot streak would be likely to continue (belief in the hot hand), but did not believe that a streak was as likely to continue when observing a non-human generating process (as found in gambling).
1.4 The Research Approach:

Our intuition tends to favor the opinions of the domain experts (professional basketball players, coaches, commentators) over the opinions of statisticians. Despite the large body of statistics provided by normative decision scientists that the hot hand is an illusion, the domain experts do believe in the validity of the hot hand. In other words, the experts believe that performance in sports is not a random process. They believe that performance is constrained – that is, it is affected by local constraints (e.g., confidence, fatigue, different opponents, etc.). Even though the exact numbers are difficult to estimate, there is no argument that the prevalence of this belief is widespread, ranging from amateurs to professionals. While the normative models may not necessarily be incorrect in their analysis with the data they used, it is possible that they might not be framing the problem appropriately. This research seeks to investigate the hot hand phenomenon from the ecological perspective.

In framing this research, it is useful to consider two properties of time series: independence and stationarity.

Independent events are characterized by a lack of memory in the system, where the previous results do not have any bearing or influence on the next result. A dependent event is one where the outcomes of previous results do have an influence on the next result, indicating there is some memory or persistence in the system. A stationary process is one where the rules governing the generator of those events are constant. For example, a constant learning curve is an example of a stationary
and dependent process – where the rules governing the learning rate does not change, even though the probabilities of success will change as one moves along the learning curve. It is possible to have a process that is stationary, yet dependent, or stationary and independent. A fair coin is stationary (the rules governing performance are constant) and independent (the outcome of the previous flips do not influence the outcome of subsequent flips). It is also important to characterize the difference between non-stationary and dependent events as well. A non-stationary process is where the rules governing performance change over time. These rules can be based on base probability (which is independent) or could be based on the outcomes of past events (which is dependent).

Figure 1 illustrates four possibilities that arise from the combination of independence and stationarity. Our research focuses on examining the different types of time series sequences that fall into one of these four Quadrants.
If events are stationary and independent, then the process is random; the only constraint on performance is a fixed probability. **Quadrant 1** is a situation in which the events are both stationary and independent (similar to conditions in a coin flip). The probability of a particular outcome (heads or tails) does not change in this case as the events go on, either by rules governing the coin, or the outcomes of previous flips. Therefore, in this case, the normative models (similar to Gilovich et
al.) would apply. If performance in sports is actually statistically independent and stationary, as in quadrant 1, then the belief in the hot hand is seen as the representativeness bias. In this case, a string of hits or misses is incorrectly seen as ‘evidence’ of dependence and the belief in the hot hand is a myth (bias) created to account for this deviation from the expectation.

**Quadrant 2** (independent but nonstationary) represents situations where the outcomes have varying probabilities depending on changing contexts (i.e., extrinsic constraints such as different opponents). This could result in varying probabilities of making a shot, but the events are still statistically independent since the probability of success is not dependent on the outcomes of previous trials, but rather due to changing external constraints. Is it possible that judgments attributed to the hot hand actually reflect sensitivity to the changing probabilities as a function of extrinsic constraints? If this were true, then there may be ecological justification for belief in the hot hand.

**Quadrant 3** represents situations where the events are stationary and dependent. In this case, there is an invariant (i.e., not dependent on context) rule governing the statistical properties of shooting, but this rule changes as a function of the shooting history. An example of this situation is a simple learning curve, where the probability of success on the next outcome is dependent on the number of previous events. The same rule (the learning curve) applies over the entire sequence, but the probabilities change as a function of that rule. This could reflect the differential results obtained with novices and experts discussed above. Novices
are still moving up the learning curve (i.e., changing probability of success), whereas experts may be on the asymptotic region of the learning curve (i.e., essentially constant performance level). Again, it might be possible that judgments attributed to the hot hand actually reflect sensitivity to the changing probabilities resulting from intrinsic constraints such as changing skill levels.

Another example of a stationary and dependent time series is one that has a performance dependency on the previous event(s) – such as a shot dependency series. If the outcome of the previous event(s) is successful, then the probability for success for the next event is higher. This situation would allow for hot or cold streaks to develop, as the performance of the previous events positively influences the next event. Judgments attributed to the hot hand might also reflect the sensitivity to the changing probabilities resulting from an intrinsic constraint such as a performance dependency.

**Quadrant 4** is a situation where the probabilities are changing due to intrinsic and extrinsic constraints resulting in a non-stationary and dependent process. Thus, there may be multiple factors impacting shooting performance. For example, learning may not depend simply on the number of previous shots, but may depend on the number of previous successes. In addition to learning, other intrinsic (e.g., confidence, fatigue) and extrinsic (e.g., different opponents) factors may impact the probability of success. If such constraints are operating, then belief in the hot hand could possibly reflect sensitivity to these constraints.
Both Quadrants 1 and 2 contain independent event sequences. In these independent events, previous results have no bearing or influence on the next results. For Quadrant 1, normative models (Gilovich et al.) attribute performance in sports to be independent and stationary, which is why a belief in a heuristic like the hot hand would be erroneous. In Quadrant 2, even though the probabilities are non-stationary, the events are independent. In this case, belief in the hot hand may be adaptive since it reflects sensitivity to the changing probabilities of success. For Quadrants 3 and 4, which contain series that are dependent (probabilities are dependent on the number of previous events), there is a definite ecological justification for the belief in the hot hand. We propose that sports performance is likely not independent and stationary (represented by Quadrant 1), and we think that real performance can be loosely modeled in one of the other Quadrants.

Based on the widespread belief that the hot hand exists in sports, as well as the previous research espousing the ecological validity of the hot hand heuristic, we are investigating this phenomenon further. There are two areas of investigation that this research focused on.

1. Is a frequency analysis an appropriate test for indexing constraints in performance data? We used Monte Carlo simulations to validate the frequency analysis method as an appropriate index of constraints in time series using rules based on constraints in each of the four quadrants in Figure 1.
2. We explored humans’ ability to discriminate a sequence constrained by the rules of each of the quadrants in Figure 1, (as validated by the frequency analysis) from a random sequence.

This research also examined the link between the detection of the constraints in the time series by the frequency analysis and the ability of humans to discriminate these constraints. The results of this link might validate the frequency analysis method as a good index of the process that people use to predict future events based on observations of past events.

In order to answer the two main questions posed by our research, simulations based on the rules of the different quadrants were generated, with results being analyzed in the frequency domain to determine the presence of pink noise (slopes less than zero and greater than -2 in log-log space). In processes that we know were constrained, the degree of slope observed would characterize the degree of dependency in the resulting series. The second area of investigation, as mentioned above, was concerned with the ability of people to detect constraints in a series of events. This research investigated whether people were able to discriminate streaky sequences from random series which were analyzed using Hierarchical Bayesian models.

This research utilized Non Linear Dynamics analysis methods. Non-Linear Dynamics assumes a closed-loop system, where there is constant feedback between the environment and the agent (iteration or self-reference). This is in contrast to the rationalist (or normative) tradition, where the models are typically based on linear
assumptions and open-loop processes. The normative literature has also reported that humans have faulty conceptions of randomness and random processes, and are unable to discern a constrained sequence from a random one. Since a constrained sequence is generated following certain rules, the recognition of those rules might be advantageous in the prediction of future events, which might lead to decisions that result in favorable individual or team performance. Additionally, normative literature in the classical tradition (Gilovich et al., 1985, etc.) relied on linear statistical models to debunk the notion of the hot hand. The statistical models used (serial correlations and runs tests) explicitly assume that events are independent of each other. Therefore, these tests examine the retention of the null hypothesis (events are independent).

The statistics used in the normative tradition assume independence between events, however, Gilden, Thornton and Mallon (1995); Van Orden, Holden and Turvey, 2003; Holden, Choi, Amazeen and van Orden, 2011 have argued that human performance is neither independent or random. Non-Linear Dynamics would reflect the inductive coupling with the environment, which is why it may be a good method for the analysis of this research.
2. MONTE CARLO SIMULATIONS

2.1. Frequency Analysis

The first step of the research is to evaluate the sensitivity of analyses based in the frequency domain for detecting the presence or absence of local structure (dependence and/or non-stationary processes). In this phase, simulated time series will be constructed with various types of internal structure – particularly to represent plausible types of constraints that might be active in the context of sporting competitions (e.g., learning). These time series will be evaluated in the frequency domain to test for deviations from a zero slope (i.e., white noise). The fundamental question is whether the frequency analysis might detect the internal constraints on processes that are missed by the traditional linear analysis methods.

The frequency analysis method can be used to examine temporal dynamics with respect to the influence of previous events on future events (or dependence between outcomes). Power spectral density correlations depict the relationship of absolute power and frequency. This is obtained by conducting an FFT (Fast Fourier Transformation) on a time series and plotting the resulting output on a logarithmic scale. When the relationship between frequency (x-axis) and power (y-axis) is plotted, a best fit line is computed (Beta slope), whose slope indicates the level of
structure (or dependence) that is present in the series. Dependence is understood to be the relationship between outcomes. A high degree of dependence suggests that the outcome of an event is heavily influenced by previous events.

The slope of the best fit line in the power spectral density plot (also known as a Beta slope) indicates the level of dependence in the series. A slope near zero is indicative of a random process or one where there is no dependency between events, and is referred to as white noise. Figure 2 illustrates the power spectral density plot for a series representing white noise.

Figure 2. Spectral plot of white noise condition.

In a white noise condition, the power is distributed equally across the range of frequencies. Note that in this condition, there is no correlation between the absolute power and frequency. If the absolute value of the slope of the power spectral density plot is significantly greater than zero then this is considered to be
evidence that the process is not random (i.e. there are dynamic constraints). A slope of -2 suggests an integral constraint, in which there is strong dependence among events (e.g., Brownian motion). Slopes greater than 0 but less than -2 suggest partial dependence (e.g. Fractals). This region is referred to as pink noise. Figure 3 illustrates the power spectral density plot for a series representing pink noise.

![Power spectral density plot for pink noise](image)

Figure 3. Spectral plot of pink noise condition

In this condition, the lower frequencies exhibit higher power, and we observe a roll off in power at the higher frequencies (accounting for the best fit line slope). The power spectral density plots generated by human behavioral data generally falls within two conditions (white and pink noise). [Gilden (1995), Van Orden, Holden and Turvey (2003)].
2.2. Method

2.2.1. Design and Procedure:

Simulations were divided up into four categories based on the four quadrants created by crossing the dimensions of independence and stationarity shown in Figure 1.

For Quadrant 1, which is the independent/stationary condition, we constructed three different probabilities of success – 0.3, 0.5 and 0.8. We ran 10 simulations with each of these probabilities. The data was then converted to intervals as will be described below and analyzed in the frequency domain in order to ascertain whether the process was indicative of white noise (or a random generating process).

For Quadrant 2, which is the independent/nonstationary condition, we set up a couple of different parameters for our simulations. The first parameter (which would include 10 simulations) was based on static probabilities of 0.2 and 0.6. The probability of alternation between these two values between each event is 0.1, which means that there is a 90% chance that the shot probability will stay at the previous value. For example, an event has a 0.6 probability of success, the next event will have a 90% chance of staying at 0.6 and a 10% chance of alternating to 0.2. Once the probability switches to 0.2, then there is a 90% chance that the probability will stay at 0.2. We also devised two additional sets of simulations, alternating with the same static probabilities of 0.2 and 0.6, but with a chance of alternation between
them of 0.25 (75% chance that the probability will persist between events) and 0.5 (50% chance that the probability persists between events).

For Quadrant 3, which is the dependent/stationary condition, we used three simple learning curves with learning rates of 0.001, 0.003 and 0.005 and starting at $P(\text{hit}) = 0$ and asymptoting at $p(\text{hit}) = 0.8$. We conducted 10 simulations with each learning rate.

In addition to the simple learning curves, we investigated the effect of performance based probabilities governing the events in the sequence. For the first sequence, the probabilities governing success of an event was dependent on the outcomes of the previous event (known as a previous 1-shot dependency). In this condition, the probability of the first event was set at 0.5, and if the first event was a hit, then the probability of the next event being successful would be 0.7. If the previous event was a miss, then the probability governing the success of the next event was lowered to 0.3. For the second set of sequences, the probabilities governing the events were based on the performance of the previous five events (known as a previous 5-shot dependency). In these sequences, the first five events were governed with a $p(\text{hit}) = 0.5$. Once the first five events were completed, the probabilities shifted depending on the performance observed. If the number of successes were between zero and two, the probability of the next event would drop to $p(\text{hit}) = 0.2$; if the number of successes were 3 out of the last five events, then the $p(\text{hit})$ remained at 0.5; whereas if the number of successes for the previous five events was either 4 or 5, then the $p(\text{hit})$ of the next event raised to 0.7. A moving
average of the previous five events was calculated to determine the performance of the previous five events which would influence the probability of success of the next event.

In summary, for Quadrant 3, we conducted two types of simulations that represent stationary dependent processes – the first being simple learning curves where the p (hit) increases as the number of events increases, and the second being performance dependencies based on the success of previous outcomes.

Quadrant 4, which represents a dependent/nonstationary condition, included simulations with a simple learning curve (k=0.002) increasing from 0.1 and asymptoting at 0.6. However, this learning curve was weighted with either a +10% or -10% in success probability with a 10% rate of alternation of these weightings, using the same logic as was described for Quadrant 2. This simulation represents the combination of intrinsic and extrinsic influences governing the probability of success of events. In these simulations, the base probability of a hit at the start was 0.1 and increased steadily, as per the learning curve (as it did in Quadrant 3’s learning curve condition), but with a 10% chance of switching the alternation to 0.001 at the next event (the 0.001 increase is from the learning curve, representing intrinsic influences, and the 10% decrease is from the switching of the weights, representing extrinsic influences, of the learning curve). We conducted 10 simulations using these parameters.

Figure 4 depicts the process used to obtain the Beta slopes (indicating the level of dependence) from the raw data.
In order to convert our sequences of hits and misses for each sequence generated into usable data for the Fourier analysis, we recorded the intervals between successes. If a sequence of hits and misses goes something like this: 1,0,0,0,1,0,1,1, we rewrite the sequence as 1,4,2,1, as the 0’s in the sequence sum together depending on the length of the interval. The interval data provides more structure in the series, as the frequency analysis is not very sensitive to binary data. The frequency domain
analysis was then conducted by taking a time series of the intervals between successes (in our case, 1024 intervals), and applying a Fourier transform on that series.

Once the interval data for each of these conditions was generated, the data was analyzed in the frequency domain to ascertain whether there is any level of dependence between events.

In addition, we conducted a Wald-Wolfowitz runs test (as used in Gilovich et al. 1985) on the same data used for the frequency analysis. This test measures the number of runs in a time series sequence (where a run is a sequence of hits or misses) compared to the number of runs that would be expected to be produced by that same series if it was governed by a random Bernoulli process. An absolute Z value greater than 1.96 (p=0.05) would indicate that the number of runs were significantly different than what would be expected by a random process, thus suggesting that the series is not random. The data from the Wald-Wolfowitz runs test was then compared with the data obtained from the frequency analysis to determine whether they were both indexing similar properties in the time series.

2.2.2. General Expectations:

We had some general expectations as to the results that we obtained.

In this experiment, we expected that the frequency analysis methods would be sensitive enough to be able to detect dependencies between outcomes of
successive shots. We expected that for the stationary independent simulations (as represented in Quadrant 1), the analysis would yield flat slopes that resemble a random process (or white noise). We expected $1/f$ (pink noise) signatures to be produced by the analysis for simulations in Quadrants 2, 3 and 4, since they all have various constraints that influence the resulting time histories. We also expected that the strength of the Beta slopes (level of dependence) would be loosely coupled to the amount of constraint applied on a sequence. We expected the lowest Beta values for Quadrant 1, followed by Quadrant 2, Quadrant 3 and Quadrant 4, respectively.

With regards to the runs test, we would expect that for Quadrant 1, the number of runs observed in our series would be similar to the expected number of runs (an absolute Z value smaller than 1.96) for that series. We would expect that for Quadrants 2, 3 and 4 (which comprised of constrained series) that the absolute Z value obtained would be greater than 1.96.

**2.3. Results**

**2.3.1. Outline**

This section reports the results obtained by examining the nature of the Monte Carlo Simulations that were simulated using the constraints of independence and stationarity as specified in the four different quadrants. The processes were simulated to obtain 1024 interval data points. A Fourier transformation was applied
and the resulting output was plotted in a log-log space. The power exponent of the output was used to determine the degree of persistence between the data points.

An outline of our general expectations is given below:

1. We expected simulations conducted in Quadrant 1 (a stationary independent process) to show no persistence between events.
2. We expected simulations conducted in Quadrant 2 (a non-stationary independent process) to show weak (greater than 0 less than 1 Beta slope values) persistence between events.
3. We expected simulations conducted in Quadrant 3 (a stationary dependent process) to show stronger (approximately 1 Beta slope values) persistence between events.
4. We expected simulations conducted in Quadrant 4 (a non-stationary dependent process) to show strong (approximately 1 or greater Beta slope values) persistence between events.

A more detailed analysis of the simulations in each quadrant is given below.

2.3.2. Independent/Stationary Processes (Quadrant 1).

Quadrant 1 represents stationary independent processes. A process is considered independent if the outcome of each event is not affected by the outcome of other events – this is sometimes referred to as a ‘random’ process. The process of flipping a fair coin is often used as a prototype of an independent process – where
the outcome of each flip is associated with a fixed probability of one-half. The term ‘stationary’ refers to the fact that the probability is constant across events (e.g., coin flips). Stationary independent sequences were generated using three different fixed probabilities (0.3, 0.5, and 0.8) over the events.

![Moving Average for sample simulations in Quadrant 1 using a window of 20 samples to compute the average.](image)

*Figure 5.* Moving Average for sample simulations in Quadrant 1 using a window of 20 samples to compute the average.

Figure 5 shows a time history of shooting performance that we simulated using these fixed probabilities. The points in the time history were computed using a running average based on a window or sample size of 20 shots. The results show that, as expected, the percentage of hits fluctuated around the probability set for each condition. Table 1 shows the grand mean and standard deviation of the conditions for \( p(\text{hit}) = 0.3 \), \( p(\text{hit}) = 0.5 \) and \( p(\text{hit}) = 0.8 \). The grand means are close to the probability of hits for each of those simulations. The Ward Wolfowitz runs test
revealed an absolute Z value of 0.45 (Table 1), within the range of the expected number of runs.

Table 1

Results of the Simulations Conducted for all Four Quadrants.

<table>
<thead>
<tr>
<th>Quadrant 1</th>
<th>Runs Test</th>
<th>Frequency Analysis Beta Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>p(hit)=0.3</td>
<td>0.296</td>
<td>0.005</td>
</tr>
<tr>
<td>p(hit)=0.5</td>
<td>0.486</td>
<td>0.011</td>
</tr>
<tr>
<td>p(hit)=0.8</td>
<td>0.805</td>
<td>0.004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 2</th>
<th>Runs Test</th>
<th>Frequency Analysis Beta Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>p(hit)=0.2 and p(hit)=0.6</td>
<td>0.377</td>
<td>0.006</td>
</tr>
<tr>
<td>10% chance of alternation</td>
<td>0.404</td>
<td>0.004</td>
</tr>
<tr>
<td>25% chance of alternation</td>
<td>0.396</td>
<td>0.003</td>
</tr>
<tr>
<td>50% chance of alternation</td>
<td>0.377</td>
<td>0.006</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 3</th>
<th>Runs Test</th>
<th>Frequency Analysis Beta Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Simple learning curves</td>
<td>0.248</td>
<td>0.005</td>
</tr>
<tr>
<td>k=0.001</td>
<td>0.248</td>
<td>0.005</td>
</tr>
<tr>
<td>k=0.003</td>
<td>0.485</td>
<td>0.004</td>
</tr>
<tr>
<td>k=0.005</td>
<td>0.606</td>
<td>0.006</td>
</tr>
<tr>
<td>Shot Dependencies</td>
<td>0.493</td>
<td>0.010</td>
</tr>
<tr>
<td>Last 1 shot dependency</td>
<td>0.396</td>
<td>0.008</td>
</tr>
<tr>
<td>Last 5 shot dependency</td>
<td>0.396</td>
<td>0.008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 4</th>
<th>Runs Test</th>
<th>Frequency Analysis Beta Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Simple learning curve and p(hit) +/-10%</td>
<td>0.354</td>
<td>0.007</td>
</tr>
<tr>
<td>k=0.002 +/-10%</td>
<td>0.354</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Note: *p≤0.05. Spectral plot slope values are based on ten simulations (total amount for each sub-condition).

As described in the methods section, the performance of each simulation was also evaluated in the frequency domain. The spectral plots obtained from simulations in Quadrant 1 were expected to yield slopes that are flat (i.e., white noise). The spectral plot slope is a good indicator of the memory or persistence that
exists in the system (that is, how the outcomes of previous events influence future events). As mentioned earlier, slopes from 0 to -2 reflect different degrees of dependence, where a 0-slope indicates no memory or dependence, and a -2-slope indicates complete dependence (an integral process). An example of a spectral plot from one of the simulations with a fixed probability of 0.5 is shown in Figure 6. As the plot shows, the slope of the line is very close to a 0-slope, which resembles a white noise spectral pattern.

We expected that the slopes for processes governed by a fixed probability would be close to 0. An analysis of 10 simulations with a fixed p (hit) of 0.5 yielded a mean slope of -0.01749 (Table 1). A one sample t-test was conducted on the values to determine whether the mean slope was different from zero. The result showed that the mean slope was not significantly different from zero \( M = 0.02, SD = 0.06 \), \( t \)
(9) = -1.05, p = 0.16]. As expected, the spectral analysis showed a pattern consistent with a white noise process in which each event was independent from the others (i.e., there was no memory or persistence).

Table 1 shows the results for two examples of biased coins [p (hit) = 0.3; p (hit) = 0.8]. Note that these would also be considered to be stationary (i.e., fixed probability), independent (i.e., no memory) processes, since the outcome of each event is based on a constant probability. The Ward Wolfowitz runs test revealed an absolute Z value of 0.65 [for p (hit) = 0.3] and 0.38 [for p (hit) = 0.8], both of which fall within the range of the expected number of runs (Table 1).

Similar to the case with a p (hit) = 0.5, the results showed a spectral pattern consistent with white noise. The mean slope for the p=0.3 condition was -0.02 and a one tailed t-test showed that this mean was not significantly different from zero [M = 0.02, SD = 0.06], t (9) = -1.05, p = 0.16]. The mean slope for the p=0.8 condition was -0.02 and a one tailed t-test also showed a non-significant difference from zero [M = 0.02, SD = 0.05], t (9) = -1.26, p = 0.11]. These results suggest that if the process is stationary and independent (as in Quadrant 1), then the frequency analysis will yield a slope that is indicative of a random process, where there is no persistence or memory between outcomes. The spectral pattern would resemble white noise regardless of the value of the p (hit). If it is a stationary independent process, where the probability of success is unchanged between events, the spectral analysis output would resemble white noise.
2.3.3. Independent/Nonstationary processes (Quadrant 2)

This condition represents a non-stationary independent process. Similar to Quadrant 1, a process is considered independent if the outcome of each event is not affected by the outcome of other events. However, the term ‘nonstationary’ refers to the fact that the probability of success is variable across events. Nonstationary independent sequences were generated using variable probability over the events. The simulations in Quadrant 2 have no memory. That is, the outcome of each event is not influenced in any way by previous outcomes. However, the generating process is variable, where the probabilities between events are changing, even though the previous outcomes do not influence the result of the next outcome.

We conducted our simulations in this condition using two different probabilities of 0.2 and 0.6. The simulations were set up with an initial probability of 0.6 with a 10% chance that the probabilities of success between events would change on a given trial.

A moving average was computed to illustrate the shooting performance of this condition, represented by Figure 7a.
Figure 7a. Moving average for sample simulation in Quadrant 2 with 10\% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples. Also shown is the shooting performance of Quadrant 1’s p (hit) = 0.5 for a reference.

Figure 7b. Moving average for sample simulation in Quadrant 2 with 25\% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples. Also shown is the shooting performance of Quadrant 1’s p (hit) = 0.5 for a reference.

Figure 7c. Moving average for sample simulation in Quadrant 2 with 25\% chance of alternation between p (hit) of 0.6 and 0.2 using a window of 20 samples. Also shown is the shooting performance of Quadrant 1’s p (hit) = 0.5 for a reference.
Figure 7a represents one of the trials in the 10% chance of alternation condition. Table 1 also shows the grand mean and the standard deviation from the same simulation for the condition with a 10% chance of alternation. The grand mean is close to the average that we would expect (0.4) since the p(hit) of this simulation varies between 0.2 and 0.6. Besides the grand mean between the two simulations on the top of the graph, the shooting performance varied with no visually discernible pattern (the performance does not generally improve or get worse as the simulation progresses). However, since there is a 10% chance of alternation between each outcome, the p(hit) would be expected to result in some streakiness in the shooting behavior, as we would expect to see some runs of shots governed by p(hit)=0.6 as well as p(hit)=0.2. The standard deviation of this simulation (Table 1) is 0.14, which is slightly higher than the control condition (quadrant 1 p(hit)=0.5) which is 0.11, which suggests a greater variation in shooting than the condition with a fixed probability. The Ward Wolfowitz runs test revealed an absolute Z value of 3.57 (Table 1), indicating that there were fewer runs than expected by a normal Bernoulli process.

In this condition, we expected to obtain spectral plots with a slope different from zero. An example of a spectral plot is of this condition [p(hit)=0.6 and p(hit)=0.2 with a 10% chance of alternation between these probabilities] is shown in Figure 8.
In this figure, the slope obtained from the plot is -0.245 which suggests a non-zero slope. An analysis of 10 simulations with a variable p (hit) alternating between 0.2 and 0.6 yielded a mean slope of -0.17 (Table 1). A one tailed t-test showed a result that was significantly different from zero [M = -0.17, SD = 0.04), t (9) = -13.43, p < 0.05].

The simulations in Quadrant 2 were set up with no dependence between successive events. However, in the simulations with a low chance of alternation (10%) between probabilities of 0.2 and 0.6, the output resulted in a non-zero slope. This non-zero slope is due to extrinsic constraints, rather than a dependency between outcomes or any other internal constraints.

In order to investigate non-stationary independent processes further, additional simulations were conducted with the same base probabilities of 0.2 and
0.6, but with 25% and 50% chances of alternation between the events. The moving averages were also computed for these conditions, and are represented in Figure 7b and 6c. In both of these conditions (25% and 50% chance of alternation), the patterns look similar to the first condition (with the alternation of 10% between the base probabilities), however, the number of alternations in the sequence increased as the chance of alternation increased (from 10% to 25% to 50%). Visually, we were unable to detect any differences in the structure of the time series. When the chance of alternation between $p(\text{hit}) = 0.6$ and $p(\text{hit}) = 0.2$ is higher, the standard deviation is smaller. The grand means for the moving averages are all around 0.4 (average of 0.6 and 0.2) suggesting that there is an equal amount of events in the sequence that are governed by $p(\text{hit}) = 0.6$ and $p(\text{hit}) = 0.2$. The Ward Wolfowitz runs test revealed an absolute Z value of 2.16 for the 25% alternation condition (Table 1), which fell outside the expected number of runs, indicating that the process might be constrained. However, the absolute Z value was 0.31 (Table 1) for the 50% alternation condition, which fell within the range of the expected number of runs.

Spectral analyses were conducted on the simulations with the 25% and 50% chance of alternation sequences, which yielded different results. An analysis of 10 simulations with a variable $p(\text{hit})$ alternating between 0.2 and 0.6 (and a chance of alternation of 25%) yielded a mean slope of -0.04 (Table 1). A one tailed t-test showed a result that was significantly different from zero [$M = -0.17, SD = 0.06$), $t(9) = -2.10, p < 0.05$]. However, an analysis of 10 simulations with a variable $p(\text{hit})$ alternating between 0.2 and 0.6 (and a chance of alternation of 50%) yielded a mean slope of 0.00199 (Table 1). A one tailed t-test showed a result that was not
significantly different from zero \([M = 0.00, SD = 0.07], t (9) = -0.08, p = 0.47\). This would make sense, as the generating process was essentially randomly shuffling between 0.6 and 0.2 with no discernible streakiness, which resulted in a white noise output for this condition.

2.3.4. Dependent/Stationary Processes (Quadrant 3)

Quadrant 3 represents a stationary dependent process. A process is considered dependent if the outcome of an event is influenced by the previous events. If a process is dependent, then it is considered a ‘non-random’ process. The process is stationary since the rule governing the process are fixed (do not change).

2.3.4.1. Learning

An example of a stationary dependent process is a simple learning curve, which is what we based a set of our simulations on. This learning curve was constructed to start at \(p\) (hit) =0 and asymptote at \(p=0.8\). We constructed this curve using a learning rate [denoted by \(k\) in the equation \(y = 0.8 \left(1 - e^{-kt}\right), k > 0\)] of 0.001. Two additional learning curves were also constructed using \(k\) of 0.003 and 0.005. The shooting performance for a simulation in this condition is depicted in Figure 9.
Figure 9. Moving averages for sample simulations of learning curves in Quadrant 3 using a window of 20 samples

In this figure, the shooting performance in question corresponds to the line denoted by \( k=0.001 \). The first several events show a lower shooting average, which increases steadily as the number of events increases. Figure 9 also shows the shooting performance simulations with different learning rates. The higher the \( k \) value, the higher the rate of learning (fewer events are needed to increase performance to an asymptotic level). The shooting performance for the curve of \( k=0.003 \) was slightly steeper than the curve of \( k=0.001 \). The curve of \( k=0.005 \) was the steepest, and reached asymptotic performance in fewer events than the other two curves. The grand means, represented for this condition in Table 1 are highest in \( k=0.005 \), followed by \( k=0.003 \) and \( k=0.001 \). This suggests that over 1000 events,
for the curve $k=0.005$, shooting performance was better than the curve $k=0.003$ and $k=0.001$. The Ward Wolfowitz runs test revealed an absolute $Z$ value of 5.19 for $k=0.001$, an absolute $Z$ value of 4.33 for $k=0.003$ and an absolute $Z$ value of 4.83 for $k=0.005$ (Table 1), indicating that there were fewer runs than expected by a normal Bernoullli process.

We expected that the learning curves in this condition would yield spectral plots that were not flat, but fall within the region of what we would classify as pink noise (or partial dependence). Since our simulations were deliberately constructed with partial dependence in mind, we would expect this dependence to be reflected in the non-zero slope of the spectral plots. An example of a spectral plot for one of the learning curve simulations ($k=0.001$) is shown in Figure 10.

![Figure 10](image)

*Figure 10.* Spectral plot of a sample simulation for learning curves in Quadrant 3 $k=0.001$. Note: in this simulation, the output slope resembles a non-zero slope.
In this condition, the output slope is -0.369, which resembles a non-zero slope, suggesting that there is partial dependence between events (pink noise). Table 1 shows the results of the simulations of the simple learning curves in this Quadrant/Condition. An analysis of 10 simulations with k=0.001 yielded a mean slope of -0.4907. A one tailed t-test indicated the slopes obtained were significantly different from zero \( \text{M} = -0.49, \text{SD} = 0.12 \), \( t(9) = -12.91, p < 0.05 \). 10 simulations of k=0.003 yielded a mean slope of -0.4521, with a one tailed t-test indicating a significant difference from zero, or white noise \( \text{M} = -0.45, \text{SD} = 0.16 \), \( t(9) = -8.89, p < 0.05 \). Also, 10 simulations of k=0.005 resulted in a mean slope of -0.4296, with a one tailed t-test indicating the slopes were significantly different from zero \( \text{M} = -0.43, \text{SD} = 0.18 \), \( t(9) = -7.55, p < 0.05 \). The slopes obtained for the three different learning curves were also not significantly different from each other. A paired 2-sample t-test between k=0.001 and k=0.003 was \( t(18) = 0.599, p \text{ (2 tailed)} = 0.56 \). Between k=0.001 and k=0.005: \( t(18) = 0.879, p \text{ (2 tailed)} = 0.39 \). Between k=0.003 and k=0.005: \( t(18) = 0.291, p \text{ (2 tailed)} = 0.77 \).

### 2.3.4.2. Local Dependence

The spectral analysis was sensitive enough to detect the dependence between events in the learning curve. The results obtained depict slopes that are within the region of pink noise, or partial dependence between events. To investigate the effect of dependence of the previous events, we conducted two sets of simulations; one where dependence was contingent on the last events, and one
where it was contingent on the previous five events. For first set of simulations (1 shot dependency), the \( p(\text{hit}) \) was initially 0.5, and if the last shot was successful, the \( p(\text{hit}) \) of the next shot was 0.7 and if the last event was unsuccessful, the \( p(\text{hit}) \) of the next shot was 0.3. For the second set of simulations (5 shot dependency), we purposely constructed a simulation where low performance on the previous five events (2 successes or less) resulted in a lower \( p(\text{hit}) \) for the next event. If the performance was higher (4 or 5 successes) then the \( p(\text{hit}) \) for the next event was also higher.

*Figure 11. Moving averages for sample simulations of shot dependency in Quadrant 3 using a window of 20 samples.*
Figure 11 is a moving average of the shooting performance for these two dependencies. Both the dependencies in this condition result in some variation in shooting performance, although the dependencies based on the last 5 shots tended to vary less. Upon visual inspection, it is not easy to pick out any discernible pattern. The grand means and standard deviations for these two moving averages are shown in Table 1. The Ward Wolfowitz runs test revealed an absolute Z value of 11.29 for the 1 shot dependency condition (Table 1), indicating that there were far fewer runs than expected by a normal Bernoulli process. Similarly, the absolute value for the 5 shot dependency revealed that the sequences exhibited fewer runs than expected by a random process (Table 1) [\(|Z| = 2.88, p < 0.05\)].

For these simulations exhibiting local dependency, our expectation was that the spectral analyses would show a slope in the region of pink noise. Since the outcome of the next event is partially dependent on the outcome of the previous event, we expected that the spectral plot would show a non-zero slope. An example of the spectral plot for the 1 shot dependency condition is shown in Figure 12.
Figure 12. Spectral plot of a sample simulation for one shot dependency in Quadrant 3. Note: This is the output from one simulation with a dependency (or memory) of one previous event. The slope of the output is close to 0, suggesting that there is no memory or dependence between outcomes, or that the analysis is not sensitive enough to detecting any dependency that exists.

This simulation shows a slope that is very close to 0 (flat) which corresponds to white noise. This could be because the frequency analysis method is not sensitive enough to detect dependence within this parameter.

The results of 10 simulations with 1 shot dependency yielded a mean slope of 0.00082 (Table 1), and a one tailed t-test indicating that these slopes are not significantly different from zero, and are thus considered to be within the region of white noise \( [M = 0.00, SD = 0.01], t(9) = 0.26, p = 0.60 \).

We assumed that since our simulations have been generated using a dependency between outcomes, and the slopes from the spectral plots are indicative of dependence, that at some point, we would be able to see a non-zero slope. We
expected that the 5 shot dependency was more likely to reveal dependence in the spectral plots. An example of the spectral plot depicting the 5 shot dependency condition is shown in Figure 13.

In Figure 13, we can see a slope of the simulation is -0.206, which is a non-zero slope. The results of 10 simulations with 5 shot dependency yielded a mean slope of -0.3124 (Table 1), and a one tailed t-test indicated these slopes to be significantly different from zero \( M = -0.31, SD = 0.07 \), \( t (9) = -14.00, p < 0.05 \). This suggests that the spectral analysis method was able to detect persistence between the events for the set of simulations containing memory for the previous five events.
However, the spectral analysis was not sensitive enough to detect persistence between outcomes for the simulations containing memory of only one previous event.

2.3.5. Dependent/NonStationary Processes (Quadrant 4).

Quadrant 4 represents a nonstationary dependent process. This is a process that has changing intrinsic and extrinsic probabilities. A process that is both dependent (such as a learning curve with changing probabilities depending on the number of shots taken) and non-stationary (such as a game-to-game situation where the probabilities are influenced by external factors) is representative of this condition.

In this condition, we have taken a simple learning curve, as we did in Quadrant 3, and added extrinsic influences on the probability of the shot, as we did in Quadrant 2, in order to conduct our simulations for this condition. The learning curve started at $p\text{ (hit)}$ of 0.1 and asymptoted at 0.6. However, we also varied the learning curve by +/-10% with a 10% chance of alternation between events. An example of the shooting performance of this condition is depicted in Figure 14.
Figure 14. Moving average for a sample simulation in Quadrant 4 using a window of 20 samples. Note: The curve is constructed with $k=0.002$ ranging from 0.1 to 0.6, with +/-0.1 with a 10% chance of alternation of +/-0.1 between events. The control curve is generated without the extrinsic constraints (+/-0.1) added to the generating probabilities.

In this figure, the shooting performance increase is compared with a simple learning curve without any variation in performance. The standard deviations of the simulations in this quadrant (simple learning curve with additional extrinsic constraints) was not higher than what was found in the simple learning curve alone (from Quadrant/Condition 3), which was not what we expected. We thought that the additional variation of the weightings would increase the overall standard deviation of the sequences. The Ward Wolfowitz runs test revealed an absolute Z value of 1.82 (Table 1), indicating that there were a similar number of runs to what we would expect in a random Bernoulli process.
The spectral plots obtained from simulations in Quadrant 4 were expected to yield slopes that were in the region of pink noise, indicative of a non-random process. Figure 15 depicts a spectral plot of one of the simulations in this condition. Figure 15 shows a slope of -0.352, appears to be within the region of pink noise.

![Figure 15. Spectral plot of a sample simulation in Quadrant 4, k=0.002 (+/-0.1). Note: The output slope is -0.352 which is a non-zero slope, suggesting partial dependence between events for this simulation.](image)

An analysis of 10 simulations with a simple learning curve ranging between p(hit) of 0.1 and 0.6, with k=0.002, coupled with +/-10% shot probability with a 10% chance of alternation, yielded a mean slope of -0.3423 (Table 1). The spectral plot indicated that the slopes were significantly different from white noise \[ M = -0.34, \text{SD} = 0.09, \text{t(9)} = -11.95, p < 0.05]
2.3.6. General conclusions:

Table 1 summarizes the results obtained across all four Quadrants. The results obtained across all four Quadrants using the spectral analysis were generally consistent with our expectations. In Quadrant 1, a stationary independent process, where the base probabilities between events are constant, the spectral plots yielded results that were not significantly different from white noise, or 0-slopes. It did not matter what the base probability was, as long as the probability was constant. For Quadrant 2, however, the spectral analysis method had mixed results. For the probabilities of 10% and 25%, the analysis did result in significantly different slopes than white noise, but for the 50% alternation between probabilities condition, there was no significant difference found. For Quadrant 3, the spectral analysis method was sensitive enough to detect dependence in the learning curves and a 5 previous event influence on the next event. However, it did not find any significant difference between the one previous event dependency. For Quadrant 4, the analysis resulted in spectral plots with a significantly different slope than zero.

2.3.7. Link Between Beta Slopes and Absolute Z values

The slopes obtained by the frequency analysis and the Z values of the runs tests seemed to be unrelated. Figure 16 depicts the relationship between the Beta slopes and the absolute Z value.
Figure 16. Relationship between Beta slopes (frequency analysis) and absolute Z value (runs test).

Figure 16 shows a weak correlation between the results of the Frequency analysis and the Wald-Wolfowitz runs test. $R(10) = 0.0536$, $p = 0.87$. This suggests that the Beta slopes are indexing different aspects of structure in the time series than the runs test.

2.4. Discussion

The spectral analysis generally was sensitive enough to detect constraints that were present in the data. We will address the results broken down by Quadrant below:
2.4.1. Quadrant 1

The spectral analysis Beta slopes showed a stationary independent process to resemble a white noise, or random process. In our analysis, we presented sequences with a base probability of 0.3, 0.5 and 0.8. The literature led us to believe that when the base probability is 0.5 (such as a fair coin), we would expect that the sequence would exhibit “maximum randomness”. In other words, the generating rules were closest to the parameters of getting alternating hits and misses. The spectral analysis turning suggesting no dependence between events is one that even lay persons would expect.

When looking at a sequence with a generating base probability that is not 0.5, such as 0.8, we would expect to see a sequence of events that would contain more hits than misses. The opposite would be true for a sequence with a base probability of 0.3, which would contain more misses than hits. The spectral analysis was able to confirm that even though the data might "appear" streaky, the generating probabilities result in a random sequence, with no dependence between outcomes.

Our expectations were confirmed, that a stationary independent process is going to produce an exponent that is flat, or within the range of white noise (and the t-test confirmed that there was a non-significant difference from a zero-slope).
2.4.2. Quadrant 2

This process was a non-stationary independent process. This means that the probabilities are changing, possibly due to some external constraints, such as different environments, changing defenses etc. This process is not dependent on internal constraints, such as previous shooting performance, or growth along a performance learning metric.

In each of the simulations conducted in this condition, the probabilities alternated between 0.2 and 0.6. However, in the first set of simulations, we expected that there would be long streaks of generating probabilities of 0.2 before alternating to 0.6 (as there was a 10% chance of alternation). Judging by the Figure 3a (in the results section), the performance generated by the sequence did appear to have more structure than the control condition (fixed probability of 50%), which was superimposed onto the graph to serve as a reference. The spectral analysis did detect that structure, as evidenced by the significant difference in the results from white noise (Table 1).

We also conducted the simulations using the same base probabilities but varying the chance of alternation to 25% between events. This would still provide some structure in the output series, as we would expect that some streakiness in the generating probabilities (although not as much as the previous set of simulations, where the alternation chance was 10% between events). We did see that there still was a significant difference from 0, although the strength of that difference varied
(for the 10% chance of alternation: $t = -13.43$, and for 25% chance of alternation, $t = -2.10$).

We observed a white noise pattern, however, when the chance of alternation increased to 50%. This means that there was a 50% chance between each event that the probability would switch from 0.2 to 0.6. If the performance time series did contain any persistence, then the spectral analysis method was unable to detect it. However, we suspect that with such a high probability of alternation between probabilities of 0.4 and 0.6, there was not enough structure in the time series to contain any persistence between events.

There is an alternative way of examining the non-stationary independent events with respect to the spectral analysis. We were under the a priori assumption that there was no memory between events in this condition. We generated the sequence to include only extrinsic probabilities, which were not dependent on any input or performance from the individual. Therefore, while the spectral analysis might be a good indicator in detecting persistence or dependence between events in a sequence (evidence of a hot hand), the analysis method might also be sensitive to other constraints that create local changes in the underlying processes governing the series.
2.4.3. Quadrant 3

In our explorations and pilot data, we conducted an analysis examining whether the spectral analysis was able to detect any persistence between events in a stationary dependent process. Therefore, we expected that the simulations that we conducted would reveal some dependence between outcomes in the spectral plots.

The simple learning curves all revealed a non-zero exponent in the spectral analysis, suggesting that there is some dependence between outcomes. This result is consistent with our expectations, since the probabilities are increasing as the sequence progresses, resulting in better performance as the sequence progresses. This change occurs across the whole sequence, which suggests that there is structure in the series that the analysis is detecting. It is interesting to see that the shallower the learning curve, the higher the mean slope for the trials. As the learning curve gets steeper, the asymptote is reached in fewer trials, and once asymptote is reached, the generating probabilities start to resemble a stationary independent process (such as a coin flip), so we would expect less structure in the time series with a steeper learning curve than a shallower one. However, a paired 2-sample t-test did not reveal any differences between the slopes obtained for the different learning curves.

In addition to simple learning curves, we examined the dependence between events when we generated rules governing probabilities based on previous performance. We analyzed two sets of simulations in this condition (of shot performance dependencies) – previous 1 shot dependencies and previous 5 shot
dependencies. Our expectation was that the spectral analysis would reveal structure in the time series, and confirm that the output that we examined was in fact streaky.

In examining the sequences generated by the previous 1 shot dependency, we were discouraged to find that the spectral analysis did not seem to detect any structure in the series that we generated – the spectral plots resembled a random process with no persistence between events.

We also generated simulations based on a previous 5 shot dependency, where the probability governing the next event was a function of the previous five events. High performance over the previous five events resulted in a high probability that the next event would be successful, whereas lower performance over the previous 5 events would result in a lower probability of success for the next event. We felt that this situation might reflect some of the dynamics that basketball players experience in a game situation; therefore we would expect that this type of sequence would reflect structure in the spectral plots (indicative of hot hand dynamics). We obtained a significant difference from a zero-slope in this case, with our spectral plots in the region of pink noise, or partial dependence. This was consistent with our expectations.

The spectral analysis did not reveal any structure in the series for the 1 shot dependency condition, even though we designed a constrained process. This suggests that for only 1 previous shot, the spectral analysis is not sensitive enough to detect the structure that might be present in that series.
2.4.4. Quadrant 4

The nature of this type of sequence would be likely to be ecologically valid, as the probability of success is influenced by both intrinsic and extrinsic factors. We added extrinsic probabilities (like we did in condition 2) to an intrinsic process (a simple learning curve, as in condition 3) to generate the rules governing the events for this condition. Both the learning curve as well as the changing probabilities (with 10% alternation between events) revealed spectral plots with exponents that were significantly different from zero, or white noise. Therefore, we expected that the sequence would have sufficient structure to reveal dependence between events. The spectral analysis confirmed our expectations.

2.4.5. Summary

In general, we feel that the spectral analysis is a good indicator of persistence or dependence between events. It was able to correctly identify a stationary independent process as white noise, where there was no correlation between outcomes. This data in Quadrant 1 also suggests that while a stationary independent process with a probability different from 0.5 might provide sequences that appear to be streaky, the generating process is unconstrained (or random) and only based on the underlying probability.

In the other conditions, where we generated the sequences to have structure, the spectral analysis seemed to reveal that the slopes were significantly different
from zero. These results warrant further examination into the appropriate way of examining sequences for streakiness or dependence. The rationalist tradition has used runs tests and serial probabilities to determine whether data appears streaky or random, and perhaps incorporating non-linear analysis methods might be useful in determining the nature of the data.

For 10 of the 12 simulations the conclusions of both tests agreed with respect to whether the process was consistent with expectations for a Bernoulli process. For the three simulations in Quadrant 1 (generated using a Bernoulli process) and one of the simulations in Quadrant 2 (in which the generating probability varied based on a 50% probability between two Bernoulli processes) both forms of evaluation showed that the resulting sequences were consistent with expectations for a Bernoulli process (runs = expectations for Bernoulli process; spectral slope = 0).

For six of the simulations both forms of evaluation found that the results of the simulation deviated from expectations for a Bernoulli process. That is, the tests correctly detected that constraints were used in the simulations. These included two of the simulations from Quadrant 2 (alternating between two Bernoulli processes with .1 and .25 probability), the 5-Shot Dependency Simulation in Quadrant 3, and the three learning simulations from Quadrant 3.

The most interesting results were where the Runs Test and the spectral analysis disagreed. For the 1-Shot Dependency Simulation in Quadrant 3, the runs test showed the strongest deviation from the expectations for a Bernoulli process.
(largest z-score). However, this constraint was not evident in the spectral pattern. 
The spectral slope was equal to zero, consistent with white noise. The opposite result was found for the single simulation in Quadrant 4 (learning + hot/cold biases). In this case, the spectral slope was significantly different than zero (suggested a constraint), while the mean number of runs was not significantly different than expected for a random process.

To further compare the two methods for evaluating the sequences, Figure 16 shows mean z-scores from the Runs Test as a function of the mean spectral slopes for each simulation. Although the tests generally agreed with respect to whether constraints were present or not, there was little correlation ($R^2 = .05$) between the size of the effects as measured by the absolute values of the mean z-scores and slopes. This might be an indication that the tests are differentially sensitive to the different dynamics that might be shaping performance. In particular, it seems that the Runs Test might be most sensitive to direct sequential dependencies (e.g., 1 Shot Dependency) and that the spectral analysis might be more sensitive to effects that are more global in impact (e.g., learning functions). In other words, the Runs Test might be more sensitive to constraints in time (one shot to the next), whereas the spectral analysis might be more sensitive to constraints over time (a broader, integral dependency on the past).

The results of our analysis are by no means representative of all types of sequences that might occur, but based on the encouraging findings in our data, we would expect that with more sophisticated modeling, we would be better able to
approximate the dynamics that exist in a natural situation. As we have speculated, the runs tests and the frequency analysis are analyzing different aspects of the time series. Future directions could involve taking existing data from a professional basketball team and finding the spectral plots (where the runs tests have traditionally failed), as it would be useful in determining whether the shooting performance of players justifies the belief in the hot hand.
3. SERIES DISCRIMINATION TASK

3.1. Introduction

The normative assumptions about streaks in sports are that they are the result of an illusory correlation between events, due to the fact that a small sample of events is incorrectly thought to be representative of a larger distribution. This is because in a large random sequence, we might observe something that appears to be “streaky”, just as in the example of coin flips, even a large random sequence can appear streaky depending on the sample examined. However, as argued previously, the ecological assumption is that the perception of streaky behavior in sports is the result of detecting constraints in the environment that influence skilled performance. The ability of a human to recognize and leverage a heuristic, such as the hot hand in sporting behavior, would be beneficial in making fast and frugal judgments that could lead to better team performance and therefore better results for a team. Lopes (1982) states that our ability to detect structure in the task environment is predicated on our ability to discriminate random from constrained data. We would like to examine, through a series of experiments, whether humans are able to discriminate constrained patterns from random patterns. This would
potentially strengthen the ecological argument that heuristics like the hot hand are not biases or errors, but adaptive smart mechanisms.

### 3.1.1. Conceptions of Randomness

The normative literature (Falk, 1981; Green, 1982; Wagenaar, 1970) states that humans are generally poor at discriminating patterns from random data. The terms gamblers fallacy and hot hand “fallacy” (frequently labeled in this way by normative researchers) suggest that humans are unable to generate or distinguish random processes from constrained processes. However, Nickerson (2002) has argued that this widely held view lacks validity. Nickerson argues that it is not possible to demonstrate conclusively whether a sequence is not a product of a random process. For example, a sequence from a fair coin of HHHHHHHH is equally likely as the sequence of HHTHTTTH, as the events are independent. Nickerson also states that although the research cannot conclusively state that people’s perception is sound, the evidence that we are poor judges of randomness is not compelling, and warrants further investigation (p. 353).

Ayton, Hunt and Wright (1989) have argued that, under the right conditions, people tend to be fairly intuitive statisticians. The strategies characterized as ‘suboptimal’ by the normative tradition (Gilovich et al.) may actually have hidden benefits when applied outside of a laboratory setting (p. 234). These benefits
concern the ability to leverage information that is present in the environment that is unaccounted for in an artificial setting. Nickerson (2002) also states that when knowledge regarding the generating process of a sequence is low, then participants would be worse at distinguishing random from constrained sequences. We could speculate that this is one of the reasons why it is difficult for researchers to grasp how experts in a domain (such as sports) can see and leverage patterns that are not necessarily detected by their statistical methods. However, if ecological notions of heuristics are valid, then the ability to discern time series with structure (streakiness) from randomly generated time series could be an essential skill for adapting to situated constraints.

3.1.2. Previous Research on Human Judgment

There have been many studies concerned with making conclusions about human’s abilities to make judgments about sequences.

Ayton and Fischer (2004) examined participant’s abilities to predict outcomes in a gambling scenario. They also experimented with binary sequences, where participants had to judge the source of the sequence (whether it represented chance performance or skilled performance). Ayton and Fischer varied the rates of alternation in the experiment. They found that participants attributed sequences with an alternation rate of 0.5 as representative of skilled performance. Ayton and Fischer attributed faulty conceptions of randomness as an explanation for these findings.
Scheibehenne, Wilke and Todd (2011) had participants predict the next symbol on a slot machine. They were presented with two sequences (one random and one correlated). Participants chose positively autocorrelated sequences more often than random sequences, but chose random sequences more often than negatively autocorrelated sequences. The authors attributed the findings to the tendency for humans to want to see patterns in data, as it is an evolutionarily adaptive mechanism.

McDonald and Newell (2009) asked participants to rate sequences based on how random they thought they were (on a scale of 1-7). They found that there was no bias towards rating sequences with high alternation rates as random with no causal information, but a bias existed when a causal mechanism for the sequences was introduced. In general, they found that people were poor judges of whether a sequence was random.

Diener and Thompson (1985) had participants judge sequences of heads and tails that were indicative of a random or nonrandom process. The results suggested that participants decide that a sequence is random by rejecting alternative nonrandom hypotheses, rather than by directly recognizing the series as representative of a random process.

In addition to making judgments about randomness using binary sequences, other studies have examined randomness in terms of examining patterns in graphical representations. Zhao and Osherson (2011) conducted a series of experiments where participants discriminated series of random from constrained
sequences. Participants examined graphical matrices to determine whether the sequences appeared random or constrained. Vandierendonck (2000) examined human’s ability to detect random time interval sequences from biased sequences through the use of audible tones. The findings showed that humans were fairly good at discriminating a random from a nonrandom time interval.

3.1.3. Our Research Approach

Our experiments utilize some of the methods used in previous research, although the sequences will be framed in a unique way that previous research has not utilized (to our knowledge). In our experimental conditions, subjects will be presented with a sequence of events, representing a high school basketball players season of shooting behavior (thus providing a context) where they are able to make a judgment as to whether the player’s shooting behavior appears random or constrained. We will present two types of sequences: the control condition (a stationary independent process), and an experimental condition (which will be generated using a combination of stationarity and dependence to produce a sequence that is representative of Quadrants 2, 3 or 4, from Figure 1.). We will ask participants to make a judgment as to what type of sequence it is. We will allow the participants to examine the long sequence as much as they want, but will restrict the number of events they can see at any particular time. We will present them with different viewing windows of 1, 4 or 16 events at a time to determine whether the task is more difficult when the cognitive load of remembering the sequence is
increased by constraining viewing window size. Lopes and Oden (1987) stated that the more information a participant is given about the task, the better they would be at discriminating it from a random process. Therefore, in our experiments, we would provide the context for how the sequence has been generated in order to give the participants the best chance for being able to discriminate a random from a constrained sequence. For example, we inform them that we will be examining player sequences and are being asked to identify players who are learning, from players who have a constant performance. The sequences from players who are learning will have different properties than players who are unconstrained. We expected that the additional information would be useful in helping participants discriminating between random and constrained sequences.

3.2. Method

3.2.1. Design

This experiment explored participants’ ability to discriminate between different types of constraints reflecting the four quadrants described earlier. The four types of discrimination were:
3.2.1.1. Independent/Stationary Processes (Quadrant 1)

The task involved distinguishing between poor and good shooters. The poor shooters (based on a simulation of 800 data points) had a shooting percentage of 0.200 and the good shooters had a shooting percentage of 0.600. A shooting percentage of 0.600 was chosen based on references from Artis Gilmore for career FG% record at 0.599. The discrimination task represents probabilities that are independent and stationary.

3.2.1.2. Independent/Nonstationary Processes (Quadrant 2)

The task was distinguishing between shooters that had changing probability (experimental condition) and a shooters with a constant probability (control condition). For the experimental condition, we constructed a simulation where the shooter had two potential shooting probabilities – 0.2 and 0.6. There was a 10% chance of alternation of the probabilities (from 0.2 to 0.6 and vice versa) between the shots, regardless of whether the shot was successful or not. For example, if the player starts out with a shooting percentage of 0.600, they have a 10% chance that the shooting percentage would change to 0.200 on the next shot. The control condition in this case was a 50% stationary shooting process.
3.2.1.3. Dependent/Stationary Processes (Quadrant 3)

This quadrant involved two different experiments.

a. The first experiment involved distinguishing between shooters that are on a learning curve (where the probability changes depending on the number of shots taken previously) and shooters that had a constant probability (control condition). In this experimental learning condition, the shooter’s shot success probability increased between 0.1 and 0.6 (k=0.003) over 800 trials. The control condition in this case, was a 50% stationary shooting process.

b. The second experiment involved discriminating between shooters who exhibit hot or cold streaks by having their shots governed by probabilities based on their previous five events. In the experimental condition, the shooters’ performance was based on the previous five shots. Similarly to the 5 shot dependency outlined in the previous section, the base probability started at 0.5 and changed after five events. If the shooter made 2 or fewer shots of the previous five attempts, the p (hit) of the next shot was 0.3. If the shooter made 3 out of the previous 5 shots, then the p (hit) of the next shot was governed by p (hit) of 0.5. If the shooter made either 4 or 5 out of the previous 5 shots, the p (hit) for the next shot was 0.7. The control condition was a sequence of shots using a stationary probability of p (hit) =0.5.
3.2.1.4. Dependent/Nonstationary (Quadrant 4)

The task involved distinguishing between shooters on a learning curve with some additional variation between games (such as environmental variables and changing defenses) and shooters that had constant probability of shot success (control condition). In the experimental condition, the shooter was on a learning curve similar to the previous condition in Quadrant 3, but with additional variability of shot success as in Condition 2. For example, a shooter would started out at 0.1 and increased to 0.6 over the 800 trials, but with an additional +/- 0.1 of variability to those probabilities (as in the independent/nonstationary condition), varying every shot with a 10% chance of alternation. The control condition in this condition was a 50% stationary shooting process.

Each of the processes represented in this phase of the experiment were replicated in the simulation condition and found to contain dependence through a Beta slope significantly different from zero.

The independent variable sample size determined how many shots could be viewed simultaneously in the viewing window. The sample sizes used were 1, 4, and 16 shots. Participants received blocks of 30 trials of each combination of the independent variables and the primary dependent variable was the percent correct discriminations within a Block of 30 trials. Each participant will view the largest window size (16) first and the smallest window size last.
3.2.2. Participants:

This experiment collected performance data on 80 participants (5 conditions x 16 participants). These participants were obtained 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.

3.2.3. Equipment:

This experiment was carried out on a standard HP Desktop PC running Windows 7. The monitor used to display the program was a Samsung 22 inch monitor. The program was developed using the Java runtime environment, and a standard optical mouse was used to control the display and select the appropriate options.

![Figure 17. Screenshot of the experimental condition program.](image-url)
Figure 17 is a screenshot of the program depicting one of the conditions that was used. The slider was placed on the top of the screen, and participants were able to scroll through all 800 events in the sequence by moving the triangular slider using the mouse. The number on the slider (595 in this case) indicated where on the sequence the viewing window started. In this Condition, which represented a viewing window size of 16, 16 events were viewed simultaneously. Once the participant decided on the nature of the sequence (in this case, whether the sequence was constrained or steady) they moved the mouse and clicked on the appropriate button. Once they made their selection, the program advanced to the next trial. A progress bar for performance in each window size was given, which provided the participant feedback about their performance.

3.2.4. Procedure:

Each participant experienced 3 blocks of 30 trials each. Window size was manipulated in a fixed order across blocks. Thus, every participant used a window size of 16 on the first block, 4 on the second block, and 1 on the third block. A trial began with the participant viewing a sequence of events through a viewing window. At the beginning of the trial, the window was positioned in the middle of the sequence of 800 events, allowing the participant to scroll through the sequence by moving the slider left or right.

Before the experiment started, the participants were briefed as to the nature of the sequence that they were experiencing (see Appendix A). For example,
in Block 1, the participant was told to choose whether they thought the sequence of events was representative of a shooter with a success rate of 20% or 60%. Once they selected the answer, the program automatically moved to the next trial. As part of the instructions, participants were able to undergo practice trials, where they were able to ask for any clarifications that they needed. Every effort was given by the experimenter to ensure that the participants understood the nature of the task they were performing. They saw a trial in each viewing window, with one trial from the experimental condition, and one trial from the control condition. Participants had the opportunity to repeat the practice trials until they understood the nature of the task. After the practice trials, they were told they would be unable to get any more assistance, and were unable to ask questions once the experimental trials started. When the experimental trials began, the program presented a block of 30 trials for the viewing window size 16. For each trial the participant was free to take as much time as necessary to scan the sequence using the slider control and make their decision.

Once the block of 30 trials was completed in window size 16, the program automatically moved to the next viewing window of 4 events. After those 30 trials, they examined the sequence with a viewing window of 1 event. The participant concluded the experiment once those blocks had been completed for a total of 90 trials (3 viewing windows X 30 trials). The time taken to complete this experiment typically ranged from 20-45 minutes.
The dependent measures collected included the proportion of correct responses (categorized in a signal detection matrix as proposed by Green and Swets, 1962), the time taken per trial, and the amount of scanning behavior. The time taken per trial, as well as the scanning behavior was collected to provide an insight into the strategies used by the participants, as well as a potential check to see if the participants were attempting to perform the task that was being asked of them. A very low average time trial, or a very small scanning variance in the trials were indications of a lack of effort by the participant, indicating that their data was not useful in understanding human ability to discriminate constrained from a random sequence. For these cases, the data was not used in the analysis.

3.2.5. Analysis

Each of the five conditions (representing the quadrants, with two in Quadrant 3) was analyzed separately. For each condition, the participants’ performance was quantified and framed into a signal detection matrix. This matrix contained the hits, misses, false alarms and correct rejections (these are explained further in the results section). From this data, each individual’s d prime (hereafter referred as d’) was calculated. d’ is essentially the sensitivity index of a statistic in units of the standard deviation of the noise distribution (which was the control condition). If all of the distributions are mostly over 0 (0 d’ level representing chance performance - e.g. 97.5% of the posterior density is above 0 for each window size), then the assumption is that people are able to detect the constrained patterns
more often that we expect them to by chance. Additionally, we were able to determine if there were any effects due to window size by examining the distributions of \( d' \) for each window size and checking the overlap of those distributions.

A hierarchical Bayesian model as applied to the theory of signal detection (as mentioned by Rouder and Lu, 2005) was used to ascertain the ability of the participants to discriminate a known streaky sequence (validated by the frequency analysis) from a random sequence of events.

There are a couple of reasons why we used a Hierarchical Bayesian model for our analysis. The first is that the analysis is hierarchical in order to account for the effects of participants and the effects of conditions, as they are repeated measures. Failure to do this might result in alpha inflation. The Bayesian analysis allows the beliefs of the performance to be bounded. If a participant obtains perfect performance (100% hit rate and 0% false alarm rate), that theoretically results in a \( d' \) that is infinite. If the \( d' \) is infinite, then the variance of \( d' \) is also infinite. The Bayesian model allows us to specify bounds for \( d' \) in order to get a functional estimate of the variability of \( d' \), which would allow us to obtain confidence intervals.

This Hierarchical Bayesian analysis was conducted using the software package Stan (2013). The specific code used to run the analyses is attached in Appendix B. The participant's performance was analyzed to investigate whether their performance exceeds that by which we may reasonably expect by chance (thus
suggesting that they are able to discern a constrained sequence from a random sequence).

3.2.6. General Expectations

Based on our intuitions about the adaptive nature of heuristics, we assumed that the participants would be able to correctly identify constrained patterns more often than not. In other words, we expected that participants would perform significantly better than chance at discriminating constrained patterns from random ones. Chance performance in this task would be 50%. Thus, if participants were able to discern constrained data from random data, then we would expect performance to significantly exceed 50%.

We also expected a significant decrement in performance as a function of the window size, with performance being better with the larger windows. This is due to participant’s ability to perceive less of the sequence at any particular time with the smaller windows, thus making the task harder. We expected that each participant would have the highest performance on window size 16 and the lowest performance on window size 1. Note, however, that learning across blocks tend to favor performance with the Window Size 1 due to the fixed order of presentation.
3.3. Results

3.3.1. Outline

This section reports the empirical results of the 80 participants measuring their ability to detect structure in a time series. The spectral analysis in the previous section of this paper suggests that the stationary independent processes do not show any persistence. In our experiment, participants examining sequences in Quadrant 1 were discriminating a good shooter from a poor shooter. They were not discriminating constrained sequences from random sequences. Participants discriminating sequences in Quadrants 2, 3 and 4 were tested on their ability to recognize a constrained sequence from a random sequence. In the conditions associated with Quadrants 2, 3 and 4, the control sequences were generated using a stationary independent process (with a \( p \text{ (hit)} = 0.5 \)) which the spectral analysis determined to resemble a random process.

The following sections are organized as a function of the Quadrants in Figure 1. These quadrants reflect different types of potential constraints that might be representative of the dynamics of performance in sports (e.g., shooting baskets).

3.3.2. Independent/Stationary Processes (Quadrant 1)

Sixteen participants undertook the experiment for Quadrant 1 which focused on distinguishing a good shooter from a poor shooter. The good shooter had a constant \( p \text{ (hit)} \) of 0.6, and the poor shooter had a constant \( p \text{ (hit)} \) of 0.2. Table 2
depicts the performance of the sixteen participants in terms of both percent correct and d'. Note that six participants correctly identified the good shooter from the poor shooter correctly for all trials in all blocks without making any errors.

Table 2

**Series Discrimination Performance (% correct and d' for each window size) in Quadrant 1**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Window Size 16</th>
<th>Window Size 4</th>
<th>Window Size 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% correct</td>
<td>d'</td>
<td>% correct</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.97</td>
<td>5.9</td>
<td>0.97</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>6.3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0.97</td>
<td>5.9</td>
<td>0.97</td>
</tr>
<tr>
<td>9</td>
<td>0.9</td>
<td>3.8</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>6.1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>6.3</td>
<td>0.97</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>6.3</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>6.2</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>6.2</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Mean 0.99 6 0.99 5.5 0.97 4.8
SD 0.03 1 0.01 0 0.03 0.8

The results indicated that for all window sizes in this condition, the performance of the subjects was significantly better than chance.
For window size 16: \[ M (d' \text{ window size 16}) = 6.00, p (d' > 0) = 0.99, 97.5\% \text{ CI (4.4, 8.1)} \].

For window size 4: \[ M (d' \text{ window size 4}) = 5.45, p (d' > 0) = 0.99, 97.5\% \text{ CI (4.3, 7.5)} \].

For window size 1: \[ M (d' \text{ window size 1}) = 4.80, p (d' > 0) = 0.99, 97.5\% \text{ CI (3.7, 6.8)} \].

Figure 18 shows the means, upper and lower boundaries of the confidence intervals for performance as a function of window size. As we can see from Table 3, pair wise comparisons for the window sizes were made for Quadrant 1. There were no significant differences in performance between the window sizes. This indicates that participants did not show higher performance in the larger window sizes.
Table 3 shows the comparisons between each of the window sizes by quadrant. Our expectation was that performance would be better (indicated by a larger d') with larger window sizes. The comparisons examined the probability that the posterior distributions over the group d' for each window size is larger for the larger window size. For example, p (d'WS16>d'WS4) examined the probability that the posterior distribution for the group d' (indicated by the mean d' value in Table 2 for the conditions in Quadrant 1) for window size 16 was larger than the group d' for window size 4.

Table 3

*Paired comparisons between the window sizes in each Quadrant.*

<table>
<thead>
<tr>
<th>Quadrant</th>
<th>Window size performance comparisons</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(d'WS16&gt;d'WS4)</td>
<td>p(d'WS16&gt;d'WS1)</td>
</tr>
<tr>
<td>Quadrant 1</td>
<td>0.6669</td>
<td>0.8362</td>
</tr>
<tr>
<td>Quadrant 2</td>
<td>0.9088</td>
<td>0.7262</td>
</tr>
<tr>
<td>Quadrant 3</td>
<td>Learning</td>
<td>0.7836</td>
</tr>
<tr>
<td></td>
<td>5 shot dependency</td>
<td>0.8938</td>
</tr>
<tr>
<td>Quadrant 4</td>
<td>0.5358</td>
<td>0.996 **</td>
</tr>
</tbody>
</table>

Note: ** indicates a significant difference in performance for an alpha of 0.05. For example: performance in window size 16 for the 5 shot dependency condition was significantly better than performance in window size 1.
For window size 16: \[ M (d' \text{ window size 16}) = 0.99, p (d' > 0) = 1, 97.5\% \text{ CI } (0.6, 1.4) \].

For window size 4: \[ M (d' \text{ window size 4}) = 0.64, p (d' > 0) = 0.9988, 97.5\% \text{ CI } (0.3, 1.0) \].

For window size 1: \[ M (d' \text{ window size 1}) = 0.90, p (d' > 0) = 0.9994, 97.5\% \text{ CI } (0.5, 1.3) \].

3.3.3. Independent/Nonstationary Processes (Quadrant 2)

Sixteen participants undertook the experiment for Condition 2 which focused on distinguishing a shooter with varying probability of success from a steady shooter with a constant \( p \text{ (hit)} \) of 0.5. The varying shooter had a \( p \text{ (hit)} \) of 0.6 or a \( p \text{ (hit)} \) of 0.2, with a 10\% chance that the sequence would alternate between the probabilities (similar to the sequence generated in the Fourier analysis section for Quadrant 2). Table 4 shows the proportion of correct responses each subject performed for each window size.
In this condition, the participants correctly identified non-stationary independent sequences (constrained) from steady \( p \) (hit) = 0.5 sequences for 68% of the trials when viewing the sequence through a window size containing 16 events. A window size of 4 and 1 saw the participants correctly distinguish between the sequences for 62% and 65% of the trials, respectively.

### Table 4.

**Series Discrimination Performance (% correct and \( d' \) for each window size) in Quadrant 2**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Window Size 16</th>
<th>Window Size 4</th>
<th>Window Size 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% correct ( d' )</td>
<td>% correct ( d' )</td>
<td>% correct ( d' )</td>
</tr>
<tr>
<td>1</td>
<td>0.6 0.7</td>
<td>0.43 -0.1</td>
<td>0.43 0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.67 0.9</td>
<td>0.43 -0.2</td>
<td>0.53 0.4</td>
</tr>
<tr>
<td>3</td>
<td>0.67 0.2</td>
<td>0.43 0.2</td>
<td>0.53 0.5</td>
</tr>
<tr>
<td>4</td>
<td>0.5 0.4</td>
<td>0.6 0.1</td>
<td>0.6 0.5</td>
</tr>
<tr>
<td>5</td>
<td>0.4 -0.1</td>
<td>0.53 0.1</td>
<td>0.43 -0.1</td>
</tr>
<tr>
<td>6</td>
<td>0.8 1.4</td>
<td>0.83 1.2</td>
<td>0.73 1.1</td>
</tr>
<tr>
<td>7</td>
<td>0.77 1.5</td>
<td>0.63 0.9</td>
<td>0.63 0.8</td>
</tr>
<tr>
<td>8</td>
<td>0.6 0.6</td>
<td>0.67 0.6</td>
<td>0.77 1.3</td>
</tr>
<tr>
<td>9</td>
<td>0.97 2.4</td>
<td>0.67 1</td>
<td>0.67 0.9</td>
</tr>
<tr>
<td>10</td>
<td>0.77 1.3</td>
<td>0.63 0.8</td>
<td>0.63 0.9</td>
</tr>
<tr>
<td>11</td>
<td>0.67 1.2</td>
<td>0.7 1.4</td>
<td>0.87 1.8</td>
</tr>
<tr>
<td>12</td>
<td>0.63 0.9</td>
<td>0.83 1.2</td>
<td>0.67 0.7</td>
</tr>
<tr>
<td>13</td>
<td>0.67 1.1</td>
<td>0.7 0.9</td>
<td>0.67 0.7</td>
</tr>
<tr>
<td>14</td>
<td>0.93 2.3</td>
<td>0.7 0.8</td>
<td>0.7 1.1</td>
</tr>
<tr>
<td>15</td>
<td>0.67 0.9</td>
<td>0.53 0.1</td>
<td>1 2.4</td>
</tr>
<tr>
<td>16</td>
<td>0.57 0.6</td>
<td>0.6 0.9</td>
<td>0.53 0.7</td>
</tr>
</tbody>
</table>

| Mean        | 0.68 1         | 0.62 0.6      | 0.65 0.9      |
| SD          | 0.15 0.2       | 0.13 0.2      | 0.15 0.2      |
Results showed that for all window sizes in this condition, the performance of the subjects was significantly better than chance, which was consistent with our expectations.

For window size 16: \[M (d’ \text{ window size 16}) = 0.99, p (d’ > 0) = 1, 97.5\% \text{ CI} (0.6, 1.4)].

For window size 4: \[M (d’ \text{ window size 4}) = 0.64, p (d’ > 0) = 0.9988, 97.5\% \text{ CI} (0.3, 1.0)].

For window size 1: \[M (d’ \text{ window size 1}) = 0.90, p (d’ > 0) = 0.9994, 97.5\% \text{ CI} (0.5, 1.3)].

The next analysis compared the results obtained by the participants to see if there were any performance decrements across the subjects when the window size decreased. Figure 19 depicts the means, upper and lower boundaries on the confidence intervals for the data obtained.
Figure 19. Group d' obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for Quadrant 2.

For the independent/nonstationary processes, the results we obtained were all inconsistent with our expectations regarding window size. There was no effect of window size between the conditions.

3.3.4. Dependent/Stationary Processes (Quadrant 3)

3.3.4.1. Learning

Sixteen participants undertook the experiment for the learning curve condition for Quadrant 3, which focused on distinguishing a constrained shooter from a steady shooter with a constant p (hit) of 0.5. The constrained sequences were governed by a stationary dependent process (a simple learning curve) starting with
a p (hit) of 0.1 and asymptoting at 0.6 (similar to the sequence generated in the Fourier analysis section for Quadrant 3). Table 5 shows the proportion of correct responses each subject performed for each window size for the simple learning curve condition (Quadrant 3).

Table 5.

*Series Discrimination Performance (% correct and d' for each window size) for the simple learning curve condition in Quadrant 3*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Window Size 16</th>
<th></th>
<th>Window Size 4</th>
<th></th>
<th>Window Size 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% correct</td>
<td>d'</td>
<td>% correct</td>
<td>d'</td>
<td>% correct</td>
</tr>
<tr>
<td>1</td>
<td>0.97</td>
<td>2.9</td>
<td>0.87</td>
<td>2.2</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>0.67</td>
<td>1.6</td>
<td>0.57</td>
<td>0.9</td>
<td>0.57</td>
</tr>
<tr>
<td>3</td>
<td>0.9</td>
<td>2.4</td>
<td>0.53</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>0.87</td>
<td>2.5</td>
<td>0.6</td>
<td>1</td>
<td>0.67</td>
</tr>
<tr>
<td>5</td>
<td>0.83</td>
<td>2.1</td>
<td>0.73</td>
<td>1.5</td>
<td>0.8</td>
</tr>
<tr>
<td>6</td>
<td>0.8</td>
<td>2.1</td>
<td>0.83</td>
<td>2.1</td>
<td>0.73</td>
</tr>
<tr>
<td>7</td>
<td>0.93</td>
<td>2.9</td>
<td>0.97</td>
<td>3.2</td>
<td>0.7</td>
</tr>
<tr>
<td>8</td>
<td>0.83</td>
<td>2.3</td>
<td>0.93</td>
<td>2.7</td>
<td>0.83</td>
</tr>
<tr>
<td>9</td>
<td>0.9</td>
<td>2.5</td>
<td>0.87</td>
<td>2.2</td>
<td>0.73</td>
</tr>
<tr>
<td>10</td>
<td>0.9</td>
<td>2.5</td>
<td>0.9</td>
<td>2.5</td>
<td>0.83</td>
</tr>
<tr>
<td>11</td>
<td>0.87</td>
<td>2.2</td>
<td>0.87</td>
<td>2.1</td>
<td>0.67</td>
</tr>
<tr>
<td>12</td>
<td>0.97</td>
<td>2.9</td>
<td>1</td>
<td>3.4</td>
<td>0.87</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2.9</td>
<td>0.93</td>
<td>3.1</td>
<td>0.83</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2.9</td>
<td>0.9</td>
<td>2.4</td>
<td>0.77</td>
</tr>
<tr>
<td>15</td>
<td>0.83</td>
<td>2.3</td>
<td>1</td>
<td>3.4</td>
<td>0.77</td>
</tr>
<tr>
<td>16</td>
<td>0.7</td>
<td>1.9</td>
<td>0.67</td>
<td>1.2</td>
<td>0.47</td>
</tr>
<tr>
<td>Mean</td>
<td>0.87</td>
<td>2.4</td>
<td>0.82</td>
<td>2.2</td>
<td>0.72</td>
</tr>
<tr>
<td>SD</td>
<td>0.1</td>
<td>0.2</td>
<td>0.15</td>
<td>0.3</td>
<td>0.14</td>
</tr>
</tbody>
</table>

In this condition, the participants correctly identified stationary dependent sequences (constrained) from steady p (hit) = 0.5 sequences for 87% of the trials.
when viewing the sequence through a window size containing 16 events. A window size of 4 and 1 saw the participants correctly distinguish between the sequences for 82% and 72% of the trials, respectively.

A hierarchical Bayesian model analysis revealed that for all window sizes in this condition, the performance of the subjects was significantly better than chance. For window size 16: \[M (d' \text{ window size 16}) = 2.44, \ p (d' > 0) = 1, \ 97.5\% \ CI (2.0, 2.9)].

For window size 4: \[M (d' \text{ window size 4}) = 2.17, \ p (d' > 0) = 1, \ 97.5\% \ CI (1.6, 2.8)].

For window size 1: \[M (d' \text{ window size 1}) = 1.28, \ p (d' > 0) = 1, \ 97.5\% \ CI (0.9, 1.7)].

Figure 20 depicts the means, upper and lower boundaries for the confidence intervals for the data as a function of window size.
Table 3 shows the pair wise comparisons for the different window sizes for the learning curve in Quadrant 3 (dependent/stationary processes). There was no significant difference between performance between window size 16 and window size 4. This means that participants did not do worse when the viewing window was reduced from 16 to 4. There was, however, a significant difference in performance between the window size of 16 and window size of 1. Participants encountered more difficulty discriminating the constrained sequence from a random sequence in window size 1 than window size 16, consistent with our expectations. Participants also found it more difficult to discriminate a constrained sequence from a random sequence in window size 1 compared to window size 4. Participants did significantly
worse in window size 1 than they did in window size 4. This was a result that was consistent with our expectations.

### 3.3.4.2 Local Dependence

Sixteen participants undertook the experiment for Condition 5 which focused on distinguishing a constrained shooter from a steady shooter with a constant $p$ (hit) of 0.5. These sequences are also part of Quadrant 3, which represents stationary dependent processes, but instead of a simple learning curve, the sequences are generated with a 5 shot dependency (similar to the Fourier analysis simulations in Quadrant 3). The sequence adjusts the probability of the next event based on the performance of the last 5 events. We expected that a sequence with these rules would generate a sequence that would appear more constrained than a stationary independent process. Table 6 shows the proportion of correct responses as a function of each window size.
In this condition, the participants correctly identified stationary dependent sequences from steady p (hit) = 0.5 sequences for 76% of the trials when viewing the sequence through a window size containing 16 events. A window size of 4 and 1 saw the participants correctly distinguish between the sequences for 71% and 69% of the trials, respectively.
A hierarchical Bayesian model analysis revealed that for all window sizes in this condition, the performance of the subjects was significantly better than chance.

For window size 16: \[ M (d' \text{ window size 16}) = 1.63, p (d' > 0) = 1, 97.5\% \text{ CI } (1.1, 2.2) \].

For window size 4: \[ M (d' \text{ window size 4}) = 1.20, p (d' > 0) = 1, 97.5\% \text{ CI } (0.8, 1.6) \].

For window size 1: \[ M (d' \text{ window size 1}) = 1.06, p (d' > 0) = 1, 97.5\% \text{ CI } (0.7, 1.4) \].

Figure 21 depicts the means and boundaries of the confidence intervals as a function of window size.

\[ Figure 21. \text{ Group d'} obtained for each window size, along with the upper (97.5\%) and lower (2.5\%) confidence intervals for the shot dependency condition in Quadrant 3. \]

Table 3 shows the pair wise comparisons for the different window sizes for the learning curve in Quadrant 3 (dependent/stationary processes). We did not
observe a significant decrement in performance between window size 16 and window size 4. We also observed no significant difference in performance between window size 4 and window size 1. There was, however, a significant difference in performance between window size 16 and window size 1, which means that the participants found it significantly more difficult to discern a constrained sequence from a random pattern with a window size of 1 than a window size of 16.

3.3.5. Dependent/Nonstationary Processes (Quadrant 4)

Sixteen participants undertook the experiment for Condition 4 which focused on distinguishing a constrained shooter from a steady shooter with a constant $p(\text{hit})$ of 0.5. The constrained sequences were governed by a stationary dependent process (a simple learning curve) starting with a $p(\text{hit})$ of 0.1 and asymptoting at 0.6, but additionally had an external weighting of +/-10% to each event’s probability, with a 10% chance of alternation of this weighting (similar to the sequence generated in the Fourier analysis section for Quadrant 4). Table 7 shows the proportion of correct responses each subject performed for each window size.
In this condition, the participants correctly identified stationary dependent sequences (constrained) from steady p (hit) = 0.5 sequences for 81% of the trials when viewing the sequence through a window size containing 16 events. A window size of 4 and 1 saw the participants correctly distinguish between the sequences for 80% and 72% of the trials, respectively.
A hierarchical Bayesian model analysis revealed that for all window sizes in this condition, the performance of the subjects was significantly better than chance.

For window size 16: \[ M (d' \text{ window size 16}) = 1.98, p (d' > 0) = 1, 97.5\% \text{ CI (1.5, 2.5})].

For window size 4: \[ M (d' \text{ window size 4}) = 1.92, p (d' > 0) = 1, 97.5\% \text{ CI (1.5, 2.5})].

For window size 1: \[ M (d' \text{ window size 1}) = 1.20, p (d' > 0) = 1, 97.5\% \text{ CI (0.9, 1.5})].

Figure 22 shows the means, upper and lower boundaries for the confidence intervals for performance as a function of window size.

*Figure 22. Group d’ obtained for each window size, along with the upper (97.5%) and lower (2.5%) confidence intervals for Quadrant 4.*
Table 3 shows the pair wise comparisons for the different window sizes for Quadrant 4 (dependent/nonstationary processes). For the dependent/nonstationary processes, there was no significant difference between performance between window size 16 and window size 4. This means that participants did not do worse when the viewing window was reduced from 16 to 4. There was, however, a significant difference in performance between the window size of 16 and window size of 1. Participants encountered more difficulty discriminating the constrained sequence from a random sequence in window size 1 than window size 16, consistent with our expectations. Participants found it more difficult to discriminate a constrained sequence from a random sequence in window size 1 compared to window size 4. This was a result that was consistent with our expectations.

3.3.6. Summary of Average Trial Time and Scanning Behavior

Our results showed that participants were able to discriminate a good shooter from a poor shooter better than chance. Our results also suggested that participants were able to discriminate a constrained sequence from a random sequence better than chance across all window sizes in all the quadrants. Additionally, we collected the average trial time, as well as the scanning behavior for all the conditions, which is shown in Table 8.
A t test revealed that participants took less time per trial in Quadrant 1 than they did in Quadrant 2. $T(4) = 6.7$, $p < 0.01$. Participants also did not take

### Table 8

**Average Trial Time and Scanning Variance for Each Window Size for Each Quadrant**

<table>
<thead>
<tr>
<th>Quadrant 1</th>
<th>Average Trial Time (secs)</th>
<th>Scanning Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size 16</td>
<td>14</td>
<td>11.71</td>
</tr>
<tr>
<td>Window Size 4</td>
<td>10.6</td>
<td>8.93</td>
</tr>
<tr>
<td>Window Size 1</td>
<td>11.5</td>
<td>10.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 2</th>
<th>Average Trial Time (secs)</th>
<th>Scanning Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size 16</td>
<td>21.5</td>
<td>33.41</td>
</tr>
<tr>
<td>Window Size 4</td>
<td>18.7</td>
<td>40.85</td>
</tr>
<tr>
<td>Window Size 1</td>
<td>16.3</td>
<td>43.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 3</th>
<th>Learning Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size 16</td>
<td>26.5</td>
</tr>
<tr>
<td>Window Size 4</td>
<td>19.6</td>
</tr>
<tr>
<td>Window Size 1</td>
<td>20.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 4</th>
<th>5 Shot Dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size 16</td>
<td>25.9</td>
</tr>
<tr>
<td>Window Size 4</td>
<td>23.7</td>
</tr>
<tr>
<td>Window Size 1</td>
<td>20.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quadrant 4</th>
<th>Average Trial Time (secs)</th>
<th>Scanning Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window Size 16</td>
<td>26.7</td>
<td>117.2</td>
</tr>
<tr>
<td>Window Size 4</td>
<td>21.3</td>
<td>135.39</td>
</tr>
<tr>
<td>Window Size 1</td>
<td>22.3</td>
<td>152.28</td>
</tr>
</tbody>
</table>
significantly more time per trial in Quadrant 4 than the simple learning curve condition in Quadrant 3, $t(4) = 2.33$, $p = 0.07$.

Interestingly, for all Quadrants, the participants more time per trial for window size 16 than window size 4: $t(4) = 4.73$, $p < 0.01$, and window size 1: $t(4) = 7.60$, $p < 0.01$. This could have been due to participants learning the nature of the task in window size 16 and thus spending more time on the trials in window size 16. There was no significant difference in trial time between window size 4 and window size 1: $t(4) = 0.62$, $p = 0.28$.

Scanning variance was lowest for Quadrant 1, compared to Quadrant 2 [$t(4) = 8.09$, $p < 0.01$], and Quadrant 3 (5 shot dependency) [$t(4) = 3.63$, $p < 0.05$]. This means that the participants did not need to scan the sequence much before they made their judgments as to whether the sequence represented a good or a poor shooter. Participants in Quadrant 4 scanned the most across all window sizes (compared to Quadrant 3 learning curves [$t(4) = 8.37$, $p < 0.01$] and 5 shot dependency [$t(4) = 11.96$, $p < 0.01$]). The learning curve condition in Quadrant 3 also saw high scanning behavior. This is because in order to determine the nature of the sequence in a learning curve, the participants need to compare performance in the beginning and the end of the sequence. In the other conditions that do not involve a learning curve, such as Quadrant 2 and the 5 shot dependency in Quadrant 3, the participants do not necessarily need to examine the beginning and the end of the sequence to determine whether it is constrained or random.
3.4. Discussion

The results indicated that across all Quadrants, the participants’ ability to detect patterns exceed those expected by chance. Across the quadrants, participants had the most success discriminating between the sequences in Quadrant 1. Quadrant 3’s simple learning curve was the next easiest task, followed by the learning curve and +/-10% weightings in Quadrant 4. The 5 shot dependency constraint (Quadrant 3) was harder than the previously mentioned conditions. Quadrant 2 proved to the most challenging to the participants. In Quadrant 1, the participants had little trouble with the task.

As mentioned above, Quadrant 1 was the easiest task. The stationary independent properties generating the sequence were very different from each other (0.6 and 0.2); hence it was relatively easy to visually see the differences in the sequences. Participants also had no trouble understanding the nature of the task, as the instructions were very simple (they did not need to understand the difference between a sequence generated by a random process and a constrained process). In this condition, since performance was so high, window size did not have a significant effect on performance.

In the other Quadrants, the participants were able to detect deviations from a stationary/ independent, or random sequence (the control condition in the conditions represented by Quadrants 2, 3 and 4) better than chance. While streakiness may be present in a stationary independent generating process, the participants looked for patterns in the constrained sequences and were more
successful than chance in determining the sequences that did not appear to be

generated by a random process.

In general, participants did not do worse when the window size reduced
from 16 to 4 (across all conditions, shown in Table 3). For the easier tasks (simple
learning curve in Quadrant 3 and Quadrant 4) window size did seem to play a role in
performance. Performance in window size 1 was significantly worse than
performance in window size 16 and window size 4. In the 5 shot dependency
condition, the only significant difference in performance was between window size
16 and 1. In the hardest condition, the window size seemed to have no effect on
performance. These results suggest that as discrimination task between a random
or constrained sequence gets more difficult, the effects of window size on
performance diminish.
4. GENERAL DISCUSSION

4.1. d’ versus Beta Slopes and Runs Tests

In this research we have attempted to provide a link between two different bodies of work. The first concept relates to spectral analysis. Much of the ecological literature (Burns, 2004; Gigerenzer, 2000) suggests that human performance is non-linear. The spectral analysis represents the effort to connect these non-linear processes to the nature of human performance.

The second body of work in the normative tradition views heuristics as erroneous. This body of work focuses on what performance should be, within the context of the linear normative models. Most of the research in this realm is context independent. Within the prescriptions of the normative linear models, the literature suggests that humans are incapable of discriminating a constrained from a random sequence. However, our experiments have grounded the simulations that we generated to the nature of sporting performance. We found that people are able to detect deviations from a random sequence, which indicates that they are more skilled than the normative models suggest.

Figure 23 links the two different areas of this research together. We have showed that the frequency analysis method is able to detect structure in a time series, and that humans are able to detect this structure in a series better than chance.
Figure 23. Relationship between frequency analysis (Beta slopes) and experimental performance (d’). Note: Quadrant 1 experiments are omitted since the nature of the task was dissimilar to the rest of the quadrants.

The data in Figure 23 suggests that the performance varies as a function of the level of dependency found by the frequency analysis, $r(2) = 0.95$, $p < 0.05$. In other words, people find it easier to discriminate a constrained sequence from a random one when the strength of constraint as measured by the spectral slope (Beta) is higher. This suggests that the Beta values in the frequency analysis may be a good index of the kind of structure that people use to predict performance of future events based on the observations of previous events (dependency). The
spectral plot slopes might have some validity to index the psychophysics of pattern detection by humans.

Figure 24 depicts the relationship the Z score obtained by the runs tests and the experimental performance of participants.

![Figure 24. Relationship between the runs test (|Z| value) and the experimental performance (d'). Note: Quadrant 1 experiments are omitted since the nature of the task was dissimilar to the rest of the quadrants.](image)

The data in Figure 24 suggests that the experimental performance is unrelated to the z score runs test values, \( r(2) = 0.0014, p = 0.99 \). In other words, the runs test is a poor predictor of human performance related to pattern discriminability. Thus, based on the results we obtained in Figures 21 and 22, the
Beta values obtained are a more powerful index of performance than the runs test, as used by Gilovich et al. (1985).

### 4.2. Implications

Our research findings serve to strengthen the claim that patterns might exist in the world that are not detected by some of the linear indexes that have been used to measure dependency (e.g. runs tests used by Gilovich et al., 1985). The normative tradition has sought to debunk human judgment with the aid of heuristics as fallacious using statistics that assume independence between events. As mentioned before, in the dominant model within the scientific community (the normative tradition), heuristics are seen as a suboptimal strategy in decision making. The reasoning behind this view is that by the normative prescriptions of rationality, heuristics (such as the hot hand) are fallacious as they assume a dependence or nonrandom process where they claim none exists. Using the statistical inference and rules of logic dominant in the normative tradition, their conclusion is that the hot hand reflects a bias due to an overestimation of patterns that make our judgment error prone. Normative literature considers the hot hand to be a fallacy, and most research literature from the normative tradition refers to it specifically as the “hot hand fallacy”.

Our experimental results suggest that:

1. Spectral analysis can be a viable method for detecting persistence in a time series.
2. Humans are able to perceive streakiness in a time series more often than not.

The spectral analysis method might be superior to a runs test or a serial correlation, as proposed by Gilovich et al. (1985) in order to determine the level of persistence in a sequence. This might be a useful process going forward to detect persistence between events in a sequence, and provide evidence for structure in a series. Our experiments also have shown that humans are able to detect these patterns more often than not, which implies that we are able to potentially recognize when someone is exhibiting performance outside of what we might expect given their career averages. The recognition that a player is currently on a hot streak has implications for performance. If a team member recognizes that a player is streaky, they are more likely to pass the ball to them to take a shot. This would ensure that the best shooter does not always bear the brunt of the burden for taking shots. The risks and rewards are more equally distributed amongst a team, and defenses would have a harder time adapting to the changing constraints of a team behaving with the belief in the hot hand. A coach who recognizes a player exhibiting the hot hand might make adjustments to their team that result in more favorable team performance as well. This strengthens the case for the ecological justification for the belief in the hot hand phenomenon.

The ecological tradition has viewed decision making and heuristics from the basis of their pragmatic value, rather than their “fit” within the bounds of statistical and logical parameters. The ecological view recognizes the coupling between the human and the environment and incorporates analysis methods that assume the
non-linearity of human performance. The very fact that the ecological community trusts the opinions of domain experts over the estimations of normative statisticians points to the ecological tendency to base decisions on abduction, rather than normative logic. Todd and Gigerenzer (2008) state that heuristics “exploit our evolved capacities”, which suggests that the decisions are based on experience. Castel, Rossi and McGillivray (2012) examined the beliefs about the hot hand across adult life spans. They mention a multitude of studies that state that there are many instances where older adults display effective judgment, reasoning and decision making. They state that “Older adults in the present sample, as well as the highly experienced fans and players tested by Gilovich et al. (1985), strongly believe in the hot hand, which may represent the greater reliance on experienced based heuristics, coupled with less emphasis placed on why events may or may not be independent” (pg. 603).

Huttel, Mack and McCarthy (2002) as mentioned in Burns (2004) state that the brain is prewired to notice streaks. They state the specific areas of the brain are activated by noticing streaks, and the strength of the activation is correlated with the length of the streak. This finding argues for an adaptive element to notice streaks. Scheibehenne et al. (2011) state that pattern recognition is an adaptive process, which is useful when we leverage limited information based in the task environment to make quick decisions.

Some research in the ecological tradition, while not even being able to validate the actual behavior of the hot hand, validate the belief in the hot hand as
advantageous. Gigerenzer and Todd (2000) discuss the criterion for evaluating heuristics is whether they produce beneficial outcomes, with respect to the attainment of their goals. In terms of the hot hand in basketball, the goal in to maximize the number of points scored to increase the chance of a win. Burns (2004) states that in a professional basketball situation, upon receiving possession, there are twenty four seconds before a shot must be taken (shot clock). Players or coaches must decide who to allocate the next shot to within a fraction of the shot clock at most so that an appropriate play may be executed. Every player may be aware of the shooting averages of their teammates, but while this is a valid allocation cue, this is not one utilized every time, otherwise the shooter with the best career average would take every single shot. Since the shooting decision is a multi-cue process, the belief in the hot hand and streakiness is potentially a valid criterion by which to allocate the shot. If the cue is valid, then it should increase scoring behavior.

Lopes (1991) has argued for the use of heuristics as a hallmark of intelligent behavior. Lopes (1981 and 1982) have also sought to argue against the normative assumptions that a simple probability is the best source of information regarding how something will perform. Lopes argues that unique situations will give rise to properties that will affect the simple probabilities that are normatively assumed to be representative, and hence these simple probabilities might be poor predictors of performance. This concept is easy to observe: A player with a 0.500 shooting average over their career will likely not exhibit the same performance in any particular game. Each game, or even fraction of a game, might have its own unique constraints, ranging from any possible combination of stationarity and dependence.
that might affect the outcome of the next shot. Perhaps it is time to abandon the normative notion of probability in decision making, and focus on the pragmatic value of a decision?

4.3. Future Directions of the Research

Our research focused on determining the structure of a time series to determine if persistence existed within the sequences that were generated. Future research might focus on obtaining actual shooting data and examining the sequence of a player over a season using the spectral analysis method described in this paper. We might also look at the level of persistence detected by spectral analysis in empirical data. Future research might focus on collecting skilled performance data using participants with varying levels of experience and examining the time series generated by their performance to examine levels of dependence. For example, would people who are on different levels of a learning curve exhibit different levels of persistence?

Nickerson (2002) states that participants perform better generating random sequences when the task is grounded with context. We grounded our discrimination task around basketball, and we found that humans were able to discriminate random from constrained sequences. Future research might focus on varying the amount of instruction and context and comparing performance to investigate effects of grounding in judgment tasks.
In our research, we found that humans are able to discriminate a constrained sequence from a random sequence. Future research might investigate where the discrimination breaks down. For example, if we generate a learning curve with similar initial and asymptotic probabilities, at what point will it become too difficult for the humans to detect constraints? We also found high performance in Quadrant 1’s discrimination task, but we used probabilities that were quite different. Future research might focus on titrating those probabilities down to investigate at what point humans are unable to detect any differences? Could we detect the difference between a sequence with a 45% base probability and a 55% probability?

Our series discrimination task involved participants judging different sequences based on the properties of dependence and stationarity as based on the four Quadrants described in Figure 1. Future research might attempt to get a sense of participant performance across the Quadrants, examining whether series discrimination is easier across some of the Quadrants. A condition might also be set up where the discrimination task focuses on two types of constrained sequences to see if participants can discriminate between them (for example, might it be easier to discriminate sequences between Quadrant 2 and 3 or Quadrants 3 and 4).

Based on our understanding of sporting behavior, we think that shooting performance is a combination of extrinsic and intrinsic constraints. We think that the generating probabilities are dependent on previous player performance (intrinsic) as well as nonstationary factors between different game situations (extrinsic). In our research of Quadrant 4 processes, which represented a
dependent/nonstationary condition, we focused on a single combination of dependence/nonstationarity, which was represented by a simple learning curve with a +/- 10% weight which had a 10% chance of alternation between events. Future research might focus on different representations of generating probabilities, such as a shot dependency (as we examined in Quadrant 3) with additional extrinsic influences (as we examined in Quadrant 2) to see if dependence would still be revealed in a spectral plot. Additionally, we could ask participants to judge those sequences to examine whether they would be able to discriminate a sequence constrained in that condition from a random sequence.

Scheibehenne et al. (2011) conducted an experiment using participants to determine random from constrained sequences and offered them the opportunity to win $50 if their performance was high. This might motivate participants more if they had an opportunity to obtain monetary compensation, and future research might incorporate a rewards system.

Previous research in the ecological tradition (e.g. Burns 2004) discusses the validity of the belief in the hot hand as leading to better overall team performance. Future research might focus on empirically researching this assertion. In a team dynamic, if a coach recognizes the hot hand (validated as a constrained sequence by frequency analysis) and leverages that information using a certain decision criterion, would that lead to better overall performance?

We determined that humans are able to discern a constrained sequence from a random sequence, which has been validated by the non-zero spectral plot slopes in
the frequency analysis. At this point, we feel that our research serves as a building block to continue the discussion on the nature of time series and human perception of constrained sequences. We have suggested through our research that spectral analysis might be a more sensitive measure than other normative measures (such as serial correlations), which accounts for the chaotic nature of human performance. We also suspect that the grounding of the context of the task (in a basketball situation for our experiment) resulted in the ability to discriminate random and constrained sequences. The correct recognition of a constrained sequence (such as the exhibition of the hot hand) has predictive value, and leveraging that information can lead to advantageous performance. If human performance is based on chaotic processes, then using heuristics (such as a belief in the hot hand) is ecologically valid and a smart mechanism, and based on our results, warrants further investigation.
APPENDIX A

Instructions for Participants for the Different Experimental Conditions

Condition 1:

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 each of the conditions 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.

In this condition, you will be attempting to identify players with a high shooting percentage (over 0.500) for the season. You will be reviewing sequences from good shooters (players with over 0.500) and poor shooters (players with averages lower than 0.500). We would like to correctly identify the good shooters so that we can try and recruit them to our program, since we currently do not have any exceptional shooters. We would like to identify shooters that can come on and make an immediate impact for our offense. Please correctly identify the good shooters (over 0.500 shooting) and poor shooters (under 0.500 shooting) from the following sequences.

*Applies to subjects 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76*
Condition 2:

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 each of the conditions 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.

In this condition, you will be attempting to identify players who are consistent shooters, versus players that are streaky shooters. The streaky shooters will appear to have strings of hits and strings of misses more often than you would expect to encounter by chance. We would like to identify streaky players so that we can take a further look into their other attributes in order to ascertain whether they will be a good fit for our team. Please correctly identify the streaky shooters from the players with a consistent shooting percentage (0.500)

*Applies to subjects 2, 7, 12, 17, 22, 27, 32, 37, 42, 47, 52, 57, 62, 67, 72, 77*
**Condition 3:**

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 each of the conditions 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.

In this condition, you will be examining players with potential, who are capable of improving as the season progresses. Part of having a good balanced team is to incorporate youngsters who will develop their skills and become better players as they gain experience at the college level. Therefore, you are asked to identify the players that show promise and learning (that is, their shot percentage improves throughout the season). Please correctly identify the players who exhibit the potential for learning from the players with a consistent shooting percentage (0.500).

*Applies to subjects 3, 8, 13, 18, 23, 28, 33, 38, 43, 48, 53, 58, 63, 68, 73, 78*
**Condition 4:**

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 each of the conditions 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.

In this condition, 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 team. Please correctly identify the streaky shooters who improve as the season progresses from the players with a consistent shooting percentage. (0.500)

*Applies to subjects 4, 9, 14, 19, 24, 29, 34, 39, 44, 49, 54, 59, 64, 69, 74, 79*
Condition 5:

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 each of the conditions 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.

In this condition, you will be examining players who are streaky. They will have different probabilities of scoring depending on how well they did in the previous 5 shots. If they are on a hot streak, they are more likely to keep scoring at a higher rate than if they are on a cold streak. These players should exhibit more streaky behavior than we would expect by chance (i.e. longer strings of hits and misses than we would expect by a shooter who is not streaky). We would like to identify these players for our program so that we can refine their shooting at the college level. Therefore, you are asked to identify the players that you think exhibit streakiness (longer strings of hits and misses than we would expect to see from a random 0.500 shooter with no streakiness).

Applies to subjects 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80.
APPENDIX B

Code used to run the Hierarchical Bayesian model in the R statistical package for the experimental data.

data {
  int<lower=0> Nsignal;
  int<lower=0> Nnoise;
  int<lower=0> Nsubjects;
  int<lower=0> Nsizes;
  int<lower=0> hits[Nsubjects,Nsizes];
  int<lower=0> fas[Nsubjects,Nsizes];
}

parameters {
  real H[Nsubjects,Nsizes];
  real F[Nsubjects,Nsizes];
  real groupH[Nsizes];
  real groupF[Nsizes];
}
real<lower=0> varH[Nsizes];

real<lower=0> varF[Nsizes];

}

transformed parameters {

real<lower=0,upper=1> phiH[Nsubjects,Nsizes];

real<lower=0,upper=1> phiF[Nsubjects,Nsizes];

real groupdprime[Nsizes];

real groupcrit[Nsizes];

real dprime[Nsubjects,Nsizes];

real crit[Nsubjects,Nsizes];

for (i in 1:Nsizes) {

groupdprime[i] <- groupH[i] - groupF[i];

groupcrit[i] <- -groupF[i];

for (sj in 1:Nsubjects) {

phiH[sj,i] <- Phi(H[sj,i]);
}
phiF[sj,i] <- Phi(F[sj,i]);

dprime[sj,i] <- H[sj,i] - F[sj,i];

crit[sj,i] <- -F[sj,i];

}

}

}

model {

for ( i in 1:Nsizes ) {

  groupH[i] ~ normal(0, 1000);

  groupF[i] ~ normal(0, 1000);

  varH[i] ~ inv_gamma(.01, .01);

  varF[i] ~ inv_gamma(.01, .01);

  for ( sj in 1:Nsubjects ) {

    H[sj,i] ~ normal(groupH[i], varH[i]);

    F[sj,i] ~ normal(groupF[i], varF[i]);

  }

}
hits[sj,i] ~ binomial(Nsignal, phiH[sj,i]);

fas[sj,i] ~ binomial(Nnoise, phiF[sj,i]);
REFERENCES


Gigerenzer, G., & Todd, P. M. (2000). *Simple heuristics that make us smart*. Oxford University Press, USA.


