**Misperceptions of Randomness: Revisiting the Hot Hand Fallacy**

*Phil166*

*Ben Peloquin*

In their seminal, 1974 work, “Judgment under Uncertainty: Heuristics and Biases,” Amos Tversky and Daniel Kahneman asked the question, “how do people assess the probability of an uncertain event or the value of an uncertain quantity?” This question is fundamental to the human experience - day-to-day life is replete with ambiguity we must somehow navigate. In perceiving the distance of an object, or assessing the intended meaning of a speaker’s utterance, we converge on some pragmatic estimate of the “truth,” despite the constraints of limited cognitive resources and often-incomplete information. In his work, “Presumptive meanings,” the Linguist Stephen Levinson provided a guiding analogy using the Rembrandt sketch shown in Figure 1:

*We interpret this sketch instantly and effortlessly as a gathering of people before a structure, probably a gateway; the people are listening to a single declaiming figure in the center. [. . . ] But all this is a miracle, for there is little detailed information in the lines or shading (such as there is). Every line is a mere suggestion [. . . ]. So here is the miracle: from a merest, sketchiest squiggle of lines, you and I converge to find adumbration of a coherent scene [. . . ].*

So how do we converge on the scene of a speaker (likely Christ) before a gateway from mere squiggles? How do we navigate everyday uncertainty in the face of cognitive constraints and ambiguity in the environment? Tversky and Kahneman (1974) argued that, for a start, we likely rely on a set of heuristic principles. Such principles, they claim, are beneficial because they reduce the complexity of the tasks we face to “simpler judgmental operations.” But these heuristics may also lead to “severe and systematic errors.”

**Current paper**

In the following paper, we review one of the more popular errors studied within this framework – the divergence of people’s intuitive beliefs of randomness from the basic laws of chance. To do so, we revisit the controversial topic of the “Hot Hand” in Basketball, first orienting the reader to the basic framework provided by Tversky & Kahneman (1974) and then turning to the original paper by Gilovich, Vallone & Tversky (1985) (GVT), which investigated the “Hot Hand phenomena” in basketball. We then review more recent work, which has called into question findings raised by the 1985 paper. Finally, we conduct a statistical analysis using a new data set of every shot taken during the 2014-2015 NBA season, re-running the analyses in Gilovich, Vallone & Tversky (1985) and assessing the shortcoming proposed by the recent work. We end with a summary of our findings.

**Tversky & Kahneman (1974)**

In “Judgment under Uncertainty: Heurisitics and Biases” Tversky and Kahneman provide a description of basic cognitive heuristics we may employ, through a series of empirical investigations. In particular, they describe three such heuristics:

1. *Representativeness*
2. *Availability*
3. *Adjustment and Anchoring*

We will focus on the heuristic of *representativeness* as it plays a role in our current work. *Representativeness* encodes the idea that in answering questions such as “What is the probability that A belongs to class B?” or “What is the probability that A originates from process B?” people often do so by relying on the *degree to which A is representative of B*. This seems intuitive enough - if A originates from process B, it likely shares similar characteristics of B. However, Tversky and Kahneman (T&K) showed that *representativeness* can lead to a number of misconceptions, some of which I enumerate here: judgments based on stereotyping, insensitivity to prior probabilities of outcomes, insensitivity to sample size, and insensitivity to predictability.

**Representativeness**

Critically, reliance on the *representativeness* heuristic can also lead to systematic *misconceptions of chance*. T&K found that people expected that sequences of events generated by a particular process should be *representative* of that process, even for short sequences. They give an example of judging the origin of a coin toss. They presented participants with three sequences of coin tosses and asked which was more likely to be produced by an ordinary fair coin:

1. H-T-H-T-T-H
2. H-H-H-T-T-T
3. H-H-H-H-T-H

Respondents systematically believed that sequence (1) was more likely to have come from the fair coin than both (2) or (3). This is surprising for two reasons. First, sequence (2) has the same number of heads and tails as sequence (1) - it has the exact same likelihood of being generated by a fair coin as sequence (1). Second, while sequence (3) contains more heads than tails and does not appear to represent the fairness of a fair coin, this is from only a small (n=6) sample of tosses. T&K took these results as evidence that people expect that the “essential characteristics of the process [which generated the sequences] will be represented not only globally in the entire sequence, but also locally in each of its parts.”

This belief in the representativeness of local sequences is unwarranted statistically and forms the basis for the well-known “Gamblers Fallacy.” People generally believe that after a long sequence of a particular outcome, such as a long sequence of a particular color in a game of roulette, the next turn is more likely to be a different color. Simply put, if there has been a long run of red, you should bet black. T&K argue that misconceptions of this sort are widespread not only in “naïve” subjects, but also in experienced research psychologists. They dubbed this belief the “law of small numbers,” which reflects the expectation that “valid hypotheses about a population will be represented by a statistically significant result,” with little regard to the size of the sample.

These systematic departures from the laws of probability are deeply rooted in our beliefs and perceptions of randomness. While tosses of a coin provide an easy characterization of chance events, we can apply the same framework to any sequence of binary outcomes. Shots of a professional basketball player, for example, where success is a basket made and failure is a miss, fit this formulation nicely. We now turn to the original study, which adopted this formulation, to study the widespread perception of the “Hot Hand” or streakiness in basketball.

**Gilovich, Vallone & Tversky (1985)**

In “The Hot Hand in Basektball: On the Misperception of Random Sequences,” Gilovich, Vallon & Tversky (GVT) investigated the widespread belief that certain NBA players occasionally perform at levels well above what would be considered normative based on historical performance. A player who is “hot” is believed to have shifted into a heightened level of performance, possibly due to increased focus, attention or motor control, or to perform better following recent success. In both circumstances, the expected outcome of this heightened ability is more runs of successful shots (streak shooting). This can be realized as longer and more frequent runs, or an increased probability of a made shot after making previous shots. Clearly, this perception of a “Hot Hand” is also consistent with the *misperceptions of random sequences* found in T&K (1975) and forms the basis of GVT’s investigation.

GVT explored the “Hot Hand” phenomena in three studies. In the first, they surveyed a population of basketball fans to assess the degree to which people believed in the idea of the “Hot Hand.” In the second study, the examined shot data from an NBA team during the 1981 season, performing three statistical analyses. Finally, GVT also conducted a controlled shooting experiment with players from an intercollegiate varsity basketball team. Given our use of NBA season data in the current study, we will focus on the first two parts of the GVT paper.

**GVT Study 1**

To assess the extent to which the “Hot Hand” phenomena existed in the socio-psychological realm of the average basketball fan, GVT recruited one hundred basketball fans from Cornell and Stanford University. *Fans* were defined as people who played basketball “occasionally” and watched at least 5 games per year. Participants were given a questionnaire, which examined their beliefs regarding sequential dependence among shots. Results indicated a widespread belief in the “Hot Hand” phenomena. 91% of participants believed that a player had a better chance of making shot after having just made two or three shots than after having missed two or three shots. 96% believed that players tend to make more shots after having made a series of shots in a row. 84% believed that a team should pass the ball to someone who has just made several shots in a row (i.e. distribute the ball to the “Hot Hand”).

GVT also asked participants to estimate the shooting percentage of a hypothetical player (with an overall 50% shooting percentage) after having just made a previous shot vs. after having just missed a previous shot. Participants believed that the shooting percentage for the hypothetical player would be higher on average (61%) after having made their previous shot, than having missed their previous shot (42%).

Clearly the perception a “Hot Hand” shooting in basketball was widespread in 1985. To test these intuitions formally, GVT conducted a statistical analysis of actual NBA shooting data, analyzing the records from 48 home games for the Philadelphia 76ers 1981 season. GVT conducted three analyses. First they examine the probability of a made shot conditioned on players’ recent history of hits and misses. Second they examined the frequency of different sequences of hits and misses in players’ shooting records. Finally, the analyzed the stability of a players’ performance across games.

**GVT Study 2**

GVT examined the conditional probabilities of nine players on the Philadelphia 76ers during the 1980-81 season. The compared the probability of making a shot conditioned on having missed the previous 1..3 shots and the probability of making a shot conditioned on having made the previous 1..3 shots. They found that for eight out of the nine players, the probability of making a shot after having made a previous shot was actually lower than after having missed the previous shot, a finding at odds with “Hot Hand” predictions.

In their second analysis, GVT conducted a Wald-Wolfowitz run test comparing the observed number of runs for each player with their expected number of runs (given their shooting percentage). Five of the nine players actually had more runs than would be expected, a finding that also ran counter to “Hot Hand” predictions (fewer runs indicates more streaks of shorter lengths).

In their third analysis, GVT examined the idea that the previous two tests may not have had sufficient power to detect “occasional ‘hot’ stretches” embedded in longer stretches of normal performance. To do this they partitioned the entire shot history for each player into non-overlapping sets of four shots. They then counted the number of sets in which performance was “high” (three or four shots made) vs “moderate” (two shots made) vs “low” (one or zero shots made), to assess the hypothesis that if a player is occasionally hot then their record would have more “high” performance sets than expected by chance. They did not see evidence for streak shooting in this analysis, performing a series of tests, none of which reached significance. They repeated this analysis four times, starting the partition of shot histories at different points, however all analyses failed to support Hot Hand predictions.

In their final analysis, GVT assessed the hypothesis that players might display more hot or cold nights, comparing their overall variability in per game shooting percentage with what would be expected on the basis of their overall history. To do so GVT compared the standard deviation of each player’s per game shooting percentage and one derived from the player’s overall shooting percentage across their season history. If the Lexis ratio of these standard errors for a given player was greater than one, then shooting percentages were fluctuating more than would be expected by chance alone. GVT did not find evidence for significant variation in shooting percentage based on this analysis, providing evidence that players did not appear to have more hot or cold nights than expected by chance alone.

From these analyses GVT concluded that there was no evidence for streak shooting in the NBA data they analyzed. Importantly, they did cite the possible confounds that accompany this type of data. For example, they were not able to account for the relative difficult of separate shots, which depend on the nature of the shot (type of shot, distance) as well as elements of the defense (how closely the shooting player is guarded), for example.

**Miller & Sanjurjo (2014)**

Findings from GVT have roiled sports aficionados, especially basketball fans, for the last 30 years. Many basketball fans and professionals continue to believe in the Hot Hand (Aharoni & Sarig, 2011; Attali, 2013; Cao, 2011) while formal studies have had difficulty identifying the phenomena, especially while using in-game basketball data. In “A Cold Shower for the Hot Hand Fallacy” (Miller & Sanjurjo, 2014) the authors avoided in-game basketball data entirely, instead conducting a controlled shooting experiment with semi-professional shooters. After orienting the reader to the experimental design in Miller & Sanjurjo (2014) and the subsequent findings we highlight the specific limitations of in-game data they describe.

**Experimental set-up**

Miller & Sanjurjo (2014) (M&S) conducted a controlled shooting experiment in two phases. In the first phase they tested whether any individual shooter in their sample showed evidence for the Hot Hand. In the second phase, they tested whether they could identify a Hot Hand out of sample.

The M&S sample consisted of semi-professional players from the Santo Domingo de Betanzos team in Spain. In total, they collected shot data from eight players from this team in both Phase 1 and 2, due to availability and time constraints. A shooting session for a player consisted of an initial warm-up phase in which the experimenter observed the player. The experimenter identified an approximate distance from which the shooter appeared to make 50 percent of his shots. (This was done to maximize the variance of shot outcomes.) This position was marked on the court. In a given session the player would take 300 shots from the marked position with a monetary insensitive for specific shots made. After every shot a trained rebounder retrieved the ball and held it from a fixed location near the basket. To signal the beginning of the next shot (at which point the rebounder would pass the ball to the shooter) the experimenter initiated a computer-generated tone. On average the next shot occurred approximately 7 seconds after the previous tone. M&S were clearly leaving nothing to chance.

Phase 2 was conducted six months later using the same protocol as in Phase 1. M&S collected additional data for one of the players who was perceived by his teammates as the “hottest” shooter, in order to maximize the power of their tests for this particular shooter.

This experimental set-up differed from the original controlled shooting study from GVT in several ways. While shooters bet before every shot in GVT, possibly leading to interaction effects of betting and shooting, shooters in M&S only shot (no betting). The incentive schemes in M&S were constant throughout the experiment while they accumulated over time and changed according the players’ bets in GVT. Shooters in M&S always shot from the same location, whereas shooters in GVT were forced to move after every shot. M&S collected 300 shots per session, while GVT collected only 100.

**Measures of interest**

Given this system for data collection M&S assessed three statistics which measured (1) how often each player was “hot” (on a streak), (2) a player’s shooting percentage conditional on having a recent streak of made shots and (3) the length of a player’s longest streak. In general M&S, took a “streak” to be three or more made shots in a row. To measure statistic (1), which they call , M&S examined the relative frequency of shots that immediately follow a streak by taking the ratio of the number of shot subsets that immediately followed a streak and the number of shots in a set of shots .

To measure whether the conditional hit rate of a player immediately following a streak of hits is better than would be expected for a player with a constant hit rate, M&S measure what they call the “hit streak momentum statistic” which was defined as the shooting percentage of the shot immediately following a streak of hits:

where is the subset of shots that immediately follow a streak of hits, is the value of the shot (1 if the shot is made).

To measure whether hit streaks are longer than would be expected if the player had a constant hit rate, M&S measure a statistic they called the “hit streak length statistic” , which simply measured the length of the longest run of hits, assessing the probability of observing a streak of this length given the assumption is normally distributed.

where is a run of hits (consecutive made shots flanked by misses) and is the collection of all runs of hits such that .

**Findings**

M&S argued that the three statistics enumerated above better captured the effects of Hot Hand shooting than in the controlled GVT experiment, through increase statistical power and improved identification. They conduct two analyses using these measures, one at the group level and one at the individual level. In particular, the authors identified one player for whom all the measures were significant, stating that the analysis of his data “demonstrates that an individual can have a substantial hot hand effect that … systematically re-occurs across time.” M&S also saw evidence for Hot Hand effects with seven of the eight players having and statistics significantly above expected values. M&S went on to re-analyze data from previous controlled experiments of this sort, including the original GVT study, finding evidence for Hot Hand shooting in individual players.

**The confounds of in-game data**

On the basis of their empirical findings, M&S argue that the Hot Hand phenomena likely exists, but that it is only detectable in highly controlled experimental domains. They suggest that their empirical findings form the basis for the widespread conception of the hot-hand and that it “likely exists in games and is detectable by decision makers.” M&S argue that any statistical analysis of in-game data is vulnerable to both false positive and false negatives without controlling for sources of variability. In particular, a shooter may appear “hot” when they have in fact just taken a string of relatively easy shots, either due to a specific offensive strategy or the particular dynamics of a particular game (a blowout in which the other teams has essentially forfeited). Likewise the effects of a “hot” shooter may be dampened by a defense that responds by guarding the player more closely, or the player attempting more difficult shots. Controlling for variables such as the quality of opponents, the fatigue or health of the player, the surrounding teammates, the league standings, the current play all represent possible sources of variation that could account for a player’s particular performance, thus potentially dampening Hot Hand effects when they exist or the inverse.

**Motivation for the current analysis**

While M&S believe that evidence from their study provides conclusive evidence for the existence of the Hot Hand phenomena, it is at odds with a fundamental component of the phenomenon – most people who believe in the Hot Hand do not observe highly controlled shooting experiments and yet they are still convinced of the phenomenon. In GVT’s first study, the authors found that the average basketball fan that believed an *arbitrary player* who shot with 50% accuracy on average should improve to 61% upon making their previous shot and decline to 42% having missed their previous shot. Clearly, this widespread belief in the brief shot histories is a different sort of phenomenon than the streakiness observed by M&S. In our current analysis, we improve on some of the shortcoming from some of GVT’s in-game data, using a novel data set of every shot taken during the NBA 2014-15 season. We restrict our analysis to those that match the initial description of the Hot Hand phenomenon.

**Summary of the data**

We scraped the entire shot history for the 2014-15 NBA season from stats.nba.com as well as player specific information for individual shooters. Table 1 includes a summary of the data set along with feature engineering.

|  |  |  |
| --- | --- | --- |
| **Player Data** | **Shot Data** | **Game Data** |
| Age | Current hit streak\* | Game ID |
| Height | Current miss streak\* | Time |
| ID | Distance X | Quarter |
| Name | Distance Y |  |
| Position | Made/Missed | - |
| Team | Time since previous shot\* | - |
| Weight | Shot Type\* | - |
| Years in league | Shot Zone | - |

Table . Data included in custom dataset. (\*) Denotes features not included in scraped data.

**Global Hot Hand**

Perhaps the most striking finding from the original GVT study was the regular and robust belief that an arbitrary basketball player will be more likely to hit a shot having made their previous. To test this global assumption we examined the probability of making a shot conditioned on the previous four shots. Figure 1 plots the conditional probability probabilities of made shots across all shots from the 2014-15 season. To examine this hypothesis simply, we conduct a two-sample t-test, comparing the probability of making a shot conditioned on a previous made shot vs a previous miss. Results indicate that the probability of making a shot after a miss is significantly greater than a make (*t(98635) = 1260.5, p < 0.001*). While this is clearly a crude measure, generalizing over all shots by all players, it makes a test of the global assumption that the probability of a made shot by an arbitrary player should increase after a make. The fact that we not only don’t detect an effect of previous making a shot, but that the direction is negative is counter to Hot Hand hypothesis and in line with data observed by GVT. Of course, this may be an effect of the one of the confounds suggested by M&S – perhaps players’ shooting percentage decreases after a made shot because they are guarded more closely or are taking harder shots. We investigate one of these (whether streak shooting predicts distance in a later analysis).

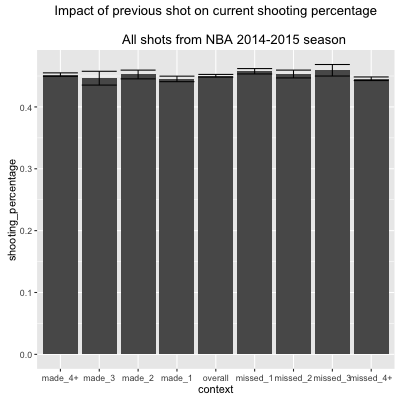


Figure . Probability of making a shot conditioned on the outcome of the previous shot across all players in the 2014-15 NBA season.

**Runs Analysis**

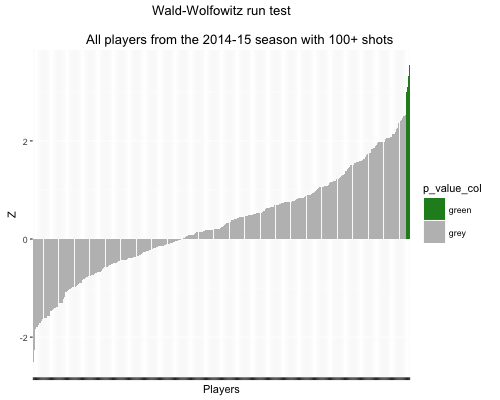
Moving from the global indications of the Hot Hand to individual players, we replicate a run-analysis GVT conducted. GVT ran Wald-Wolfowitz run tests for ach player in their dataset. In this test, each sequence of hits or misses is counted as a run. The more a player’s hits and misses cluster together, the fewer runs they have. GVT did not find any evidence for significant runs. We repeat this test, analyzing all players who have taken over 100 shots in the season. This left us with a sample of 353 players from the 2014-15 season. Figure 2 plots Z-scores for each player’s run test. If a player has significantly *less* runs than expected we should expect *negative* Z-scores (observed runs – expected runs). Results indicate that only four players had significant Z-scores, however these scores were all positive indicating that the player’s actually had more runs than was expected by chance.

Figure . Wald-Wolfowitz run tests for all players from the 2014-15 NBA season who took at least 100 shots. Green coloring denotes a significant p-value (p < 0.001). Players with significantly more runs than expected by chance should display negative scores.

In our current analysis we do not attempt to replicate findings in GVT or in M&S. Rather, using a novel data set of every shot taken in the 2014-15 NBA season, we explore a number of the potential confounds of looking at in-game data and assess the impact of streak (Hot Hand) shooting on the probability of making a basket. For the purpose of this project and given the size of our data set, we adopt a league-wide perspective, only looking at data across all shooters. Findings from the original GVT study that over 90% of Basketball fans believed that an *arbitrary* player has a better chance of hitting a shot given a previously made shot support this perspective.

**Current data**

They argued that and can detect the Hot Hand in players regardless of the frequency at which they become hot (a player who only occasionally gets hot or frequently get hot) while also accommodating periods in which the same player alternates between hits and misses when not streaking. The statistic provides greater power as its null distribution is generated by the entire sequence of a player’s shots – an improvement over a conditional

**Summary of the data**

We scraped the entire shot history for the 2014-15 NBA season from stats.nba.com. See Table 1 for a description of the data.

In the following paper we examine popular example of how these

Outlnie

1. Review of Hot Hand paper / Tversk & Kahneman
   1. What are these heuristics (Kahnemna & Tversky)
   2. What is the hot hand (Gilovich, Vallone, Tversky)
      1. Findings
2. Review of work questioning initial analysis
   1. The Hot-hand Fallacy: Cognitive Mistakes or Equilibrium Adjustments? Evidence from MLB (Green, Zwiebel)
   2. A cold shower for the Hot Hand Fallacy
3. Introduce new data set / specific hypotheses we’re interested in
4. Analysis
5. General discussion