A Cold Shower for the Hot Hand *Fallacy*

Joshua B. Miller[a] and Adam Sanjurjo[b] [*†‡§]

This Version: December 15, 2014
Previous Version: June 13, 2014

Abstract

The hot hand fallacy has long been considered a massive and widespread cognitive illusion with important economic consequences. While the canonical domain of the fallacy is basketball, which continues to provide its strongest and most readily generalizable supporting evidence, the fallacy has been considered as a candidate explanation for various economic and financial anomalies. We find, in its canonical domain, that the belief in the hot hand is not a fallacy, and that, surprisingly, the original evidence supports this conclusion. Our approach is to design a controlled shooting field experiment and develop statistical measures that together have superior identifying power over previous studies. We find substantial evidence of the hot hand, both in our study and in all extant controlled shooting studies, including the seminal study (which found the opposite result, and coined the term "the hot hand fallacy"). Also, we observe the hot hand effect to be heterogeneous across shooters, which suggests that decision makers (e.g. players and coaches) may have incentive to identify which shooters have a greater tendency to become hot. Accordingly, we find evidence that expert players (teammates) can do so. In light of these results, we reconsider the economic relevance of the hot hand fallacy more generally.

JEL Classification Numbers: C12; C14; C91; C93; D03.
Keywords: Hot Hand Fallacy; Hot Hand Effect.

---

[*]a: Department of Decision Sciences and IGIER, Bocconi University, b: Fundamentos del Análisis Económico, Universidad de Alicante. Financial support from the Department of Decision Sciences at Bocconi University, the Spanish Ministerio de Ciencia y Tecnología and Feder Funds (SEJ-2007-62656), and the Spanish Ministry of Economics and Competition (ECO2012-34928) is gratefully acknowledged. We thank José Valeiro, president of the Pabellón Municipal de Betanzos, for providing exclusive use of the basketball court, Javier Lopez for his jack-of-all-trades assistance in organizing and conducting experimental sessions, and Cristina Lopez for greatly improving the translation of our experiment's instructions to Spanish.

[†]Both authors contributed equally, and alternate order of surnames across joint papers.

[‡]This draft has benefitted from helpful comments and suggestions and we would like to thank Pedro Albarran, Jim Albert, Alberto Alesina, Jose Apesteguia, David Arathorn, Jeremy Arkes, Olivier Armantier, Daniel Benjamin, Alberto Bisin, Marco Bonetti, Gary Charness, Lucas Coffman, Vincent Crawford, Martin Dufwenberg, David Eil, Gigi Foster, Andrew Gelman, Nicola Gennaioli, Uri Gneezy, Matt Goldman, Robin Hogarth, Kyle Hyndman, Richard Jagacinski, Elliot Ludvig, Daniel Martin, Daniel Miller, Raymond Nickerson, Daniel Oppenheimer, Raffaella Piccarreta, Giovanni Ponti, Justin Rao, Colin Raymond, Aldo Rustichini, Angel Sanjurjo, Andrei Shleifer, Connan Snider, Joel Sobel, Charles Sprenger, Hal Stern, Neil Stewart, Daniel Stone, Sigrid Suetens, Richard Thaler, Robert Wardrop, Thomas Youle, and Jeffrey Zwiebel. We would also like to thank participants at the SITE Psychology and Economics Workshop at Stanford University (2014), the SABE conference (2014), the North American Econometric Society Meetings (2014), the CAGE Conference, Warwick (2014), the ESA meetings in Honolulu, Santa Cruz and Zurich (2013-2014), and the BEELab conference in Florence (2013), as well as seminar participants at Pompeu Fabra University, Simon Fraser University, Bocconi University and the University of Alicante. All mistakes and omissions remain our own.

[§]We thank Thomas Gilovich for locating the paper records of his data and scanning them for us, as well as for making us aware of the work of Jagacinski, Newell, and Isaac (1979). We thank Richard Jagacinski for locating his computer punch card print-outs and having them converted to electronic format for us.

A particular test can detect only a particular pattern or class of patterns, and complete randomness can therefore only be disproved, not proved. (Houthakker 1961)

1 Introduction

An individual believes in the hot hand if he or she thinks good outcomes have a tendency to cluster, either because a good outcome is perceived as more likely than usual after a streak of good outcomes, or because streaks of good outcomes appear as unusually long or frequent. This belief becomes a fallacy in an environment in which good outcomes arrive *as if* they have a constant rate of occurrence.

The *hot hand fallacy* was introduced in the seminal study of Thomas Gilovich, Robert Vallone and Amos Tversky (1985; henceforth GVT), in which they demonstrated that while observers, players, and coaches of basketball believe in a "hot hand" in shooting performance, the data fails to support these beliefs. In particular, the authors used evidence from in-game performance and a controlled shooting experiment to demonstrate that observed sequences of shot outcomes were statistically indistinguishable from repeated realizations of a properly weighted coin, i.e. iid Bernoulli trials (Gilovich, Vallone, and Tversky 1985; Tversky and Gilovich 1989a,b).

The GVT study was a novel exhibit of how existing laboratory results on the systematic human misperception of randomness in sequential data (Tversky and Kahneman 1974; Wagenaar 1972)—results whose general relevance had been challenged (Einhorn and Hogarth 1981; Lopes 1982; Morrison and Ordeshook 1975)—could be demonstrated in a natural setting.[1] GVT were also the first to identify a setting in which biased beliefs were both costly and acted upon by *experts* in their domain of expertise.

Over the last three decades, the hot hand fallacy has earned the reputation of "a massive and

[1]The existence of a phenomenon known as the *gambler's fallacy* in the laboratory "predicted" the presence of the hot hand fallacy in natural settings (Bar-Hillel and Wagenaar 1991). These early laboratory studies demonstrated that when people are told that a sequence of outcomes is generated by repeated coin flips, they expect outcomes to alternate more than they actually do, and for the relative frequency of an outcome in a finite sequence to closely match its underlying probability of occurrence. These behaviors are thought to lead to the gambler's fallacy, in which people attach undue significance to streaks which happen to be typical of a random device, and strengthen their belief that a streak will end, the longer the streak. Further, when people are unsure about how data is generated (as in basketball), streaks that are typical of a random device lead people to incorrectly reject the possibility that the data is random, presumably because they instead believe that there is some persistent mechanism behind the streak. See Rabin (2002) and Rabin and Vayanos (2010) for a model consistent with this behavioral process. For alternative approaches, see Barberis, Shleifer, and Vishny (1998) and Mullainathan (2002).

widespread cognitive illusion" (Kahneman 2011), owing to the sustained divide between basketball professionals' persistent and costly belief in the hot hand (Aharoni and Sarig 2011; Attali 2013; Bocskocsky, Ezekowitz, and Stein 2014; Cao 2011; Neiman and Loewenstein 2011; Rao 2009a), and the repeated inability of formal studies to discover a hot hand effect.[2] While it is not unheard of for professionals to make systematic mistakes, including in the domain of sport, the hot hand fallacy is exceptional because professionals typically respond to incentives, learn from experience, and correct their mistakes when they are made aware of them (Hakes and Sauer 2006; List 2003).[3]

Because the strength and persistence of this apparent irrationality is so striking, and the perception of patterns in sequential data is important in many domains of economic decision making, the hot hand fallacy has been given considerable weight as a candidate explanation for various puzzles and behavioral anomalies identified in the domains of financial markets (Barberis and Thaler 2003; De Bondt 1993; De Long, Shleifer, Summers, and Waldmann 1991; Kahneman and Riepe 1998; Loh and Warachka 2012; Malkiel 2011; Rabin and Vayanos 2010), sports wagering (Arkes 2011; Avery and Chevalier 1999; Brown and Sauer 1993; Camerer 1989; Durham, Hertzel, and Martin 2005; Lee and Smith 2002; Paul and Weinbach 2005; Sinkey and Logan 2013), casino gambling (Croson and Sundali 2005; Narayanan and Manchanda 2012; Smith, Levere, and Kurtzman 2009; Sundali and Croson 2006; Xu and Harvey 2014), and lotteries (Galbo-Jørgensen, Suetens, and Tyran 2013; Guryan and Kearney 2008; Yuan, Sun, and Siu 2014).

As the central contribution of this paper we demonstrate, by means of new evidence and a novel empirical strategy, that in the canonical domain of basketball, where the supporting evidence for the hot hand fallacy has long been considered the strongest, the belief in the hot hand is not a fallacy. Our approach is to conduct a controlled shooting field experiment (with expert shooters) that improves on previous designs, introduce statistical tests with greater identifying power, collect all available incentivized controlled shooting data (including that from GVT's seminal study), and conduct our improved analysis on each dataset. In contrast with previous results, we find strong

[2] Thaler and Sunstein (2008) summarized the hot hand literature in basketball with the statement: "Many researchers have been so sure that the original Gilovich results were wrong that they set out to find the hot hand. To date, no one has found it." More recent studies of in-game data present clear evidence that players do not shoot with a constant rate of success, even when controlling for shot difficulty (Arkes 2010; Bocskocsky et al. 2014; Yaari and Eisenmann 2011). While these studies have limitations (see Section 2.1), the performance variation they document is consistent with the possibility of substantial hot hand effects, and contrast with the results from the original GVT study.

[3] Notably, in the domain of sport, Hakes and Sauer (2006) confirmed the existence of systematic and costly expert mistakes in the sport of professional baseball (the *Moneyball* anomaly [Lewis 2004]), and then demonstrated that experts corrected these mistakes as they became aware of them.

evidence of hot hand shooting on both the individual and aggregate level in each dataset—consistent with the beliefs of experts and fans alike.[4] Also, we observe that the degree of hot hand shooting varies considerably across shooters (some are even the opposite of hot), which suggests that decision makers (e.g. teammates and coaches) may have incentive to correctly identify which shooters have a greater tendency to become hot. Accordingly, we find evidence that expert players (teammates) can do so.

In light of our findings, in Section 5 we reconsider the relevance of the hot hand fallacy for economic decision making more generally. Upon assessing the related literature, we find that while many studies provide evidence that is consistent with the hot hand fallacy, the evidence is surprisingly limited in its ability to identify beliefs as being both fallacious and strongly held. In particular, the body of existing evidence does not support the notion that the hot hand fallacy is a powerful and economically meaningful cognitive illusion—the current consensus drawn from evidence in the basketball domain.[5] Thus, the present finding that the hot hand effect exists in the domain of basketball, and in the original data of GVT, is central to the assessment of the relevance of the hot hand fallacy more generally, and in particular, to its robustness (or lack thereof) to expertise, incentives, and learning.[6,7]

In Section 2.1 we detail why in-game data has serious limitations as a source of evidence for (or against) the existence of hot hand shooting. Many of the limitations are standard, have already been pointed out by others, and have to do with the presence of confounding variables that are unrelated to a player's shooting ability, yet correlated with performance both on the current shot, and on the shots that precede it. In basketball game data we argue that these variables cannot be

[4] The test procedure that we outline in Section 3.2 and apply in Section 4 was not practically feasible at the time GVT conducted their analysis, given limitations in computational power.

[5] Presuming that there is no hot hand effect, there is abundant evidence that hot hand beliefs have a costly impact on decision making, e.g. "hot" players take more difficult shots, coaches allocate more playing time to "hot" players, and opposing teams allocate more defenders to "hot" players (Aharoni and Sarig 2011; Attali 2013; Bocskocsky et al. 2014; Cao 2011; Rao 2009a).

[6] In fact, GVT were careful to make this point and limit their conclusion that there is no hot hand effect to the domain of basketball because the hot hand fallacy was the general phenomenon of interest, and the evidence for the associated mistaken beliefs had been validated (Gilovich et al. 1985; Tversky and Gilovich 1989b).

[7] There are many studies that make important contributions to the study of streaks and momentum in human performance, but they have had little influence on the hot hand fallacy literature because the beliefs and behavior of decision makers have not been identified (Alter and Oppenheimer 2006). A large literature exists on the presence or absence of the hot hand effect in other sports besides basketball, for a review, see Bar-Eli, Avugos, and Raab (2006), for a meta-analysis, see Avugos, Köppen, Czienskowski, Raab, and Bar-Eli (2013b). In the finance literature there is mixed evidence of a hot hand effect (or "performance persistence") among fund managers (Carhart 1997; Hendricks, Patel, and Zeckhauser 1993; Jagannathan, Malakhov, and Novikov 2010).

sufficiently controlled for, and we demonstrate how insufficient control can allow a hot hand to go undetected, or even lead to a spurious attribution of clustering in performance to a hot hand effect. Though these limitations are particularly severe in GVT's original study of in-game data, GVT also conducted a controlled shooting experiment which addressed many of these limitations. While it seems clear that a controlled shooting experiment is necessary, we push the point further by detailing why it is likely *impossible* to identify or rule out a hot hand effect with in-game shooting data alone, despite the richness of modern data sets. As a result, in-game shooting data is unlikely to provide more than suggestive evidence for or against a hot hand fallacy in the basketball domain.

In contrast to in-game shooting, a controlled shooting experiment with expert players, similar to the one conducted by GVT, can provide the opportunity to convincingly identify, or rule out, the hot hand effect. In Section 2.2 we describe the field experiment that we conducted with a group of Spanish semi-professional basketball players.[8] The experiment improves on earlier controlled shooting tasks by removing design features that introduced confounds, such as players betting on their own shots and changing locations after each shot. In addition, we collect three times as many shots in each shooting session when compared to the original study of GVT, increasing the power of statistical tests.[9] Most importantly, we conduct multiple shooting sessions with each player, six months apart, and allocate additional sessions to the first session's "hottest" shooter, which not only further increases the power of tests for the existence of the hot hand effect on the individual and aggregate levels, but also allows us to test whether the effect can be reliably predicted out-of-sample.

In Section 3 we describe our empirical strategy, and the advantages it has with respect to GVT and others. In choosing our test statistics our objective was to accommodate the mechanisms most commonly associated with hot hand shooting: (1) a shooter shifting to a relatively higher performance state (due to, for example, internal or external factors leading to increased focus, attention, or motor control [Churchland, Afshar, and Shenoy 2006; Csikszentmihalyi 1988; Kahneman 1973]), (2) a shooter performing better following recent success (due to, for example, positive feedback lead-

[8] In the taxonomy of Harrison and List (2004), this is a *framed field experiment*.

[9] As an illustration of this increased power, when testing whether a player's shooting percentage increases immediately following three or more successful shots in a row, our design is expected to generate more than 30 observations per session for testing purposes, whereas GVT's design (a single session study) is expected to generate only 12 observations.

ing to increased confidence [Bandura 1982]), or some combination of (1) and (2).[10] Under either of these mechanisms one should expect increased shooting ability to translate into streak shooting, i.e. runs of consecutive hits (successful shots) should be longer and occur more frequently, and a player's hit rate should be higher immediately following a streak, than what would be expected if his shooting ability were instead fixed.[11] We introduce a novel package of statistical measures that not only capture these precise patterns in streak shooting, but also have improved power over the statistics employed in previous studies (see Appendix B.2 for a discussion of power). Further, we demonstrate that our measures correspond more closely to common conceptions of hot hand shooting than those employed by GVT, in part because we isolate patterns associated with "hot hand" shooting in ways previous studies do not. These issues are important because while previous studies have failed to find evidence of a hot hand effect, this may simply be due to their relative lack of power, or alternatively, to the statistics employed measuring patterns that can actually obscure hot hand shooting.[12] Finally, our test procedure can also be applied to shooting data from earlier controlled shooting studies, including the data from all extant studies pooled together.

In Section 4 we present our results. We find strong evidence of hot hand shooting at the individual level, and with substantial effect sizes—equivalent to the difference between the median and the best professional shooter.[13] In testing for the existence of the hot hand effect on the individual level we place special focus on two expert shooters—one in our sample of eight, "RC," who participated in six 300-shot sessions, and another, "JNI6," from the study of six shooters by Jagacinski et al. (1979; henceforth JNI), who participated in nine 60-shot sessions. That these shooters each shot in a large number of sessions allowed for a rigorous test for the existence of the hot hand in each shooter. Further, because RC shot in two phases, six months apart, we were able to test whether the hot hand can be predicted out of sample. We find that it can be; RC shows strong evidence of the hot hand in both phases, and his hot hand appears in nearly all of his

[10] Also see Gilovich et al. (1985) and Tversky and Gilovich (1989a) for a description. The hot hand has also been referred to as being in the *zone*, in *rhythm*, or in a state of *flow* (Csikszentmihalyi 1988), which is often described as a state of exceptional concentration and motor control.

[11] For a discussion of streaks see Gilovich et al. (1985); Gould (1989); Tversky and Gilovich (1989a); Wardrop (1999).

[12] For example, a shooter who performs better than usual immediately following streaks of three or more hits, and equally better than usual immediately following streaks of three or more misses, would (wrongly) appear to have a constant hit rate if only the difference between these two conditional hit rates were considered, as opposed to also considering differences with conditional hit rates for shots taken immediately following exactly 2 hits in a row and no more, exactly 1 hit in a row and no more, etc.

[13] Data from 2013-2014 NBA three-point percentages (NBA.com). The National Basketball Association is North America's professional basketball league.

sessions. Similarly, for JNI6, we find that his hot hand also systematically appears in nearly all of his nine sessions. In addition, we re-analyze the GVT dataset using our hit streak statistics, and find that the number of players with performance patterns in the direction predicted by the 'hot hand' hypothesis is significantly greater than would be expected if there were no hot hand effect, and that the size of this effect is substantial.

Beyond the strong evidence of individual level hot hand effects is the unexpected finding of an average hot hand effect in a pooled analysis of shooters from all extant controlled shooting studies. Because basketball is a team sport with a high degree of specialization, there is substantial heterogeneity among shooters, so one should expect heterogeneity in the hot hand effect. In fact, this is precisely what is found in controlled shooting studies—many shooters actually perform worse after recent success, and therefore an average hot hand effect appears unlikely to emerge when pooling their shooting data with that of hot players. Nevertheless, we find that the hot hand exists as an average effect in our panel of shooters. In particular, the effect is first found in phase one of our study, then this result correctly predicts the effect also being found in phase two of the study, six months later. Perhaps more remarkably, when using our statistical measures to analyze the GVT data, which has long been considered all but conclusive evidence against the existence of the hot hand, we find that the hot hand is an average effect there as well. Finally, after also finding the hot hand in JNI's data, we pool all extant controlled shooting data—ours, JNI's, and GVT's—and find the hot hand to be an average effect in a pooled sample of all shooters from all extant controlled shooting experiments.[14]

Our body of results establish that, contrary to the conclusion of hot hand fallacy drawn in the previous literature, but consistent with the ubiquitous belief among professional players, coaches, and lay-people alike, substantial hot hand shooting exists, and even appears to be a general property of the average shooter in all extant controlled shooting tasks, including the original GVT study. A natural question then follows: do these results constitute evidence of a hot hand effect in games that can be detected by expert decision makers? In our view, the evidence suggests that we should not remain agnostic. For one, that we find the hot hand effect to be robust across all extant

[14] Only one other controlled shooting experiment, Avugos, Bar-Eli, Ritov, and Sher (2013a), has been cited in the hot hand literature. Their design was a close replication of GVT's, but with fewer shots per session (40). The authors declined to make their data available. Wardrop (1999) provides a case study involving a single shooter, but after personal communication with the shooter who conducted the study herself, we viewed it as not having sufficient control to be included in our analysis.

controlled shooting designs, which are all different, suggests that it would also be present in other settings that involve the same physical act. Second, our expert subjects' beliefs about the their teammates' tendency to exhibit hot hand shooting are shown in Section 4.4 to be remarkably well-correlated with the hot hand shooting we observe in our experiment, despite the fact that these beliefs are based only on previous playing experience. That players' perceptions of hot hand shooting of teammates in games and other uncontrolled shooting environments can be used to accurately predict which teammates have a greater tendency to exhibit hot hand shooting in our controlled shooting experiment suggests that shooters who tend to be (more) hot in games also tend to be (more) hot in our controlled shooting task, and visa versa. In addition, while in-game data is subject to the aforementioned limitations, recent pooled analyses of in-game data, which have found a significant positive serial dependence in shooting performance (Arkes 2010; Bocskocsky et al. 2014; Yaari and Eisenmann 2011), are suggestive of hot hand shooting; though effect sizes in these studies appear modest, their aggregate nature may also mask substantial individual hot hand effects due to the infrequency of hot hand shooting (Arkes 2013) and heterogeneity in the size and sign of the effect among players. Finally, in a game environment, coaches and teammates, who know each other well, have access to considerably more information than whether a "hot" player's previous shots simply hit or missed; in conjunction with this, they observe how cleanly the ball entered the rim (or how badly it missed), the player's shooting mechanics, ball trajectory, body language, and other signals of the player's underlying mental and physical state, which can help them distinguish between a lucky streak of hits and a hot hand.[15]

The implications of our results extend far beyond the domain of basketball, and present a serious challenge to the widely accepted notion of the hot hand fallacy as a massive and widespread cognitive illusion, robust to the influence of experience, incentives, and learning; we elaborate this argument in Section 5.2. Further, in Section 5.3 we consider the portability of our results to a broader class of performance environments, and the implications, more generally, on the study of momentum in human performance.

[15] Bill Russell, in his 1969 retirement letter, after 13 years as a player and 3 years as a player/coach (with 11 championships), wrote: "But nobody keeps statistics on other important things—the good fake you make that helps your teammate score; ... the way you recognize when one of your teammates has a hot hand that night and you give up your own shot so he can take it" (Russell 1969).

2 Design

We conduct a field experiment to test for the existence of the hot hand effect in basketball shooting. We first motivate why a controlled shooting experiment is necessary, then describe the setting and design of our experiment, and finally highlight novel design features with respect to the existing literature.

2.1 Why Controlled Shooting?

Revisiting the hot hand fallacy in the domain in which the evidence has been considered the strongest necessitates the discussion of issues that are specific to the collection and analysis of basketball shooting data. In this section we detail why in-game data cannot provide conclusive evidence for or against the hot hand effect, and why a controlled shooting study is necessary.

If a careful researcher could control the conditions of a basketball game in any way desired, with the purpose of generating the best possible data with which to test for the existence of hot hand shooting, this researcher would observe expert shooters, shooting many times from the same location, with identical defensive pressure, with the same score in the game, surrounded by the same group of players, and with all other sources of shooting performance variability unrelated to hot hand shooting also held constant.

If this researcher does not control sufficiently for other sources of shooting performance variability, then the hot hand cannot be cleanly identified in game data, as any statistical test becomes vulnerable to false positives and false negatives. In the data a shooter may appear to have a period of hot hand shooting, but this may be because he is in the midst of a sequence of relatively easy shots, or because he shoots relatively better when the offensive strategy allows him to anticipate his shooting opportunities, the score of the game is lopsided, etc. Likewise, a shooting sequence may appear as if it were generated by a process no different than repeated coin flips but actually be generated by a shooter who, when hot, takes more difficult shots, is guarded by superior defenders, or passes the ball away relatively more often to take advantage of additional defensive pressure sent his way, etc. These concerns are not merely theoretical; recent studies have shown that when one does not attempt to control for shot difficulty, shooting performance exhibits a slight negative serial dependence (Bocskocsky et al. 2014; Rao 2009a), a phenomenon which should be expected if the opposing team makes strategic adjustments in response to a basketball player having the hot hand (Dixit and Nalebuff 1991, pg. 7).[16]

While the difficulty of controlling for the above mentioned *within* game confounds makes identification of the hot hand effect seem impossible, there are also *across* game confounds to account for, which all but guarantee that it is. In particular, because individual shooters do not shoot frequently enough in a single game for statistical tests to have sufficient power to detect abnormally streaky shooting, in-game shooting data must be pooled across a large number of games (Arkes 2010; Bocskocsky et al. 2014; Rao 2009a; Yaari and Eisenmann 2011).[17,18] By pooling shooting data across games one introduces a selection bias that can easily lead to a clustering of shot outcomes that appears to be due to a hot hand, when in fact, it is not; this bias can arise when periods of better-than-usual shooting performance are "selected" due to variables that are impossible to simultaneously control for, such as variation in the quality of opponents, player health, which teammates are playing, standings, contract situations, shot technique, offensive and defensive strategies, etc. For example, one can consider free-throw data, in which players have been found to shoot, on average, around 2 percentage points higher after hitting their first shot than after missing it (Arkes 2010; Yaari and Eisenmann 2011).[19] When analyzing the seven most recent seasons of professional basketball (NBA) free throw data, we find that free throw performance varies substantially over the course of a season. Namely, players have a 2 percentage point higher success rate in the second half of the season ($p < .0001$), meaning that if a player has hit the first shot, it is more likely that his shot is selected from the second half of the season.[20] Further, when controlling for the possibility of fixed games effects due to omitted variables, we find that the performance after a hit is not significantly better than after a miss in six out of seven seasons, with a difference of

[16] See Aharoni and Sarig (2011) and Green and Zwiebel (2013) for an expanded discussion of strategic confounding in game data.

[17] Assuming the existence of hot players, this sample size issue can be exacerbated by fewer shooting opportunities for hot players due to the strategic adjustments made by the defense.

[18] The low frequency with which shots are taken raises another issue not discussed here. Suppose a player becomes "hot" and hits a shot, but there are not enough recent shots to associate with it (or none at all); how can one detect the hot hand in this case? Without a method of measuring factors that are more tightly related to a player's hot state—such as the quality of a player's shooting mechanics, current physiological state (e.g. neural noise [Churchland et al. 2006]), or current psychological state (Bandura 1982)—detection of hot hand shooting in games is likely impossible, despite the richness of modern shooting data (e.g. STATS, Inc.'s SportVU optical tracking data).

[19] Similar pooling issues present themselves when analyzing data from the NBA's three-point contest, but in this case, because data is so sparse, it must be pooled across years (Koehler and Conley 2003). The three-point contest has further issues arising from the fact that shot locations differ and performance incentives are non-constant.

[20] Goldman and Rao (2014) document another factor that leads to between game variation in performance: on average players have a higher free throw percentage at home, but in pressure situations this reverses.

one percentage point or less in most seasons (though the difference is significant when all seasons are pooled together).[21] Yaari and Eisenmann (2011) and Bocskocsky et al. (2014) control only for within-game sources of variation that are unrelated to hot hand shooting, and therefore their results are similarly vulnerable to omitted variables that may lead to across game differences in performance.

Furthermore, once one begins controlling for within game confounds, let alone across game confounds, shooting data becomes too sparse to allow statistical testing on the individual shooter level, due to insufficient power. Instead, shooting data must be pooled not only across games, but also across players (Arkes 2010; Bocskocsky et al. 2014; Yaari and Eisenmann 2011). Even if one is willing to ignore the aforementioned limitations, when shooting data is pooled across players, identification of the hot hand is confounded by player heterogeneity; some players may perform relatively better after recent failure than after recent success, and therefore pooling their shooting data with hot players will dampen the estimated hot hand effect, and could even reverse it.[22]

Notice that if the shooting environment in a game were sufficiently controlled for a hot hand to be identifiable, then it would no longer be a game at all, but rather a controlled shooting task. Therefore, we conduct a controlled shooting experiment.[23]

2.2 Controlled shooting field experiment

Our design consists of two phases, conducted six months apart: Phase One tests whether any individual shooter in our sample has the hot hand, and whether the hot hand is an average effect in our pooled sample of shooters. Phase Two tests whether the hot hand effect can be predicted out of sample. To this end we had players from Phase One return for multiple additional sessions, to see if those with the hot hand in Phase One also have the hot hand in Phase Two, and whether any average hot hand effect in Phase One would re-occur. We first describe the setting and design, then highlight novel features with respect to previous work.

[21] Game-level controls do not influence the 2005-2006 estimate of Arkes (2010). We collected data from the 2005-2006 season to the 2011-2012 season, and extend the analysis conducted in Arkes (2010) (using a linear probability model). We find that the effect size reported in Arkes maintains when all seven seasons are pooled together, without fixed game effects.

[22] We find evidence of this effect in our data. There are players that perform better after a streak of misses than after a streak of hits, presumably because they make some sort of adjustment after a streak of failures.

[23] In Section 5.3 we discuss what we can infer about hot hand shooting in uncontrolled game environments on the basis of our controlled shooting results.

Setting and participants

We recruited players from the semi-professional basketball team Santo Domingo de Betanzos, in the Spanish province of Galicia, by dropping by at the end of a team practice and inviting all players to participate in a scientific study of basketball shooting with financial incentives.[24] While player interest was unanimous, it was not possible to accommodate all players given their limited time availability and our limited set of available time-slots. In total, eight players were able to participate in both phases of our panel.[25] The players averaged 24 years of age, and 14 years of experience playing in competitive, organized, basketball leagues. The experiment was conducted on their home court, the Pabellón Municipal Polideportivo de Betanzos, where they both practice and host opposing teams in league games. All shooting sessions were video-recorded.

Design of the shooting session

Upon arrival at the scheduled time the shooter (subject) was given several minutes to warm up by shooting however he liked. The experimenter observed the shooter in order to gauge from what distance he would make around 50 percent of his shots (in order to maximize the variance of shot outcomes for the purpose of statistical testing). The experimenter then used a strip of masking tape to mark the shooting location from which that player would take all 300 shots.[26] Next, the shooter was led to a closed room, where the experimenter read the instructions aloud as the shooter read silently (see Appendix A.1 for the shooter's instructions). The shooter was informed that he would be taking 300 shots, and that in addition to a 5 Euro participation fee, 10 of these shots would be selected at random to determine his payoffs. For the 10 randomly selected shots, he would receive 6 Euros for each shot that he hit and 0 Euros for each shot that he missed. He was also informed that the 10 shots had already been selected, printed on a sheet of paper, and sealed in an envelope. The envelope was shown to the shooter and left in his field of vision for the duration of the session.

[24] In Spain there are five rated categories of organized, competitive basketball. The top level is the ACB, in which the top earners make millions of euros each year. The second level is also professional, in the sense that players make enough money not to need other forms of employment to live comfortably. Levels three through five are considered semi-professional in the sense that while players have all of their basketball-related expenses paid for them, and may earn some take-home earnings on top of this, it is typically not enough to live comfortably on without additional employment. Santo Domingo de Betanzos this year is the best team in the 5th category, for their region, so can move up to the 4th category next year if they elect to.

[25] One of these players had recently left Santo Domingo to play professionally in the 2nd category (see previous footnote), but continued to train frequently with Santo Domingo.

[26] The shooting location was kept constant for the purpose of controlling a player's probability of hitting a shot. While the distance from the rim selected for each shooter varied, all selected locations were straight in front of the rim, meaning that they were situated on the imaginary axis which bisects the width of the court.

Upon completing the instructions the shooter was given an opportunity to ask questions before returning to the court, where he was then allowed two minutes of practice shots from the marked location before beginning the paid task.

After the shooter had an opportunity to warm up, each of the 300 shot attempts went as follows: a trained rebounder, who was unaware of the purpose of the study, held the ball from a fixed location near the basket, which was also marked on the floor.[27] When the experimenter initiated a computer-generated tone, the rebounder passed the ball to the shooter in a precisely prescribed way (see Appendix A). Once the shooter received the ball, the rebounder turned his back to the shooter and the shooter was allowed to choose the timing of his shot without constraint, though the shooters typically shot within 1-2 seconds after receiving the ball. After the shot, the rebounder collected the ball, returned to his marked location as quickly as possible, and awaited the same computer-generated tone to signal the beginning of the next shot attempt, which occurred, on average, 7 seconds after the previous tone. The task continued in this way, with the experimenter calling out after each block of 50 shots was completed. The duration of each 300 shot session was approximately 35 minutes, which was calibrated to avoid fatigue effects.[28]

Phase One and Phase Two

In Phase One each of ten shooters participated in a single session consisting of 300 shots. In Phase Two, six months later, each shooter participated in multiple additional shooting sessions. Eight of the ten Phase One shooters were available to participate in Phase Two; we refer to these eight shooters as the *panel*.[29] At the conclusion of Phase One subjects filled out a short questionnaire. Then, following Phase Two they filled out a more detailed questionnaire, which we discuss in Section 4.4.

Before Phase Two, we conducted a statistical analysis of the shooting data from Phase One

[27] The rebounder was informed only of his own task, and that the shooter would shoot 300 times.

[28] In order to minimize the possibility of a potential fatigue effect from over-exertion we settled on 300 shots after running pilots with ex-basketball players that were not in basketball shape. These players reported no problem shooting 300 times under the conditions of our design (with a rebounder). It is safe to say that less than one quarter of each session was spent in physical movement for the shooters. In a post-experiment questionnaire our subjects reported below average levels of fatigue. Players commented that they shoot around the same number of shots on a daily or near daily basis. In Section 4.1 we find that there is no evidence of fatigue effects (or warm-up effects) in our data.

[29] One of the two shooters that did not participate in Phase Two canceled at the last minute (the other was out of the country), so three additional players who were aware of the experiment and eager to participate contacted us and were given the canceled time slots. While these players were not part of our two-phase design, we nevertheless include their shots on our pooled analysis in Section 4.3.

(see Section 4.2), which identified one of our shooters, "RC," as the hottest shooter in the sample. Further, in questionnaire responses, his teammates—who had not observed RC's shooting session, but averaged 800 hours of previous playing experience with him—ranked RC as by far the hottest shooter of the group. On this basis, in Phase Two we allocated a larger number of our scarce session timeslots to RC (5) than to the other shooters in the panel (3 each), in order to maximize the power of our test of whether RC had the hot hand, and in order to observe if the effect is persistent across many sessions. Phase Two sessions were conducted using a design and protocol identical to those of Phase One.

Discussion of our shooting design

Our controlled shooting design improves on GVT's in several ways: (1) our shooters are strategically isolated; they only shoot, whereas GVT's were forced to bet before each shot, which allowed for possible interaction effects between shooting and betting,[30] (2) our shooters have constant, non-negligible performance incentives, whereas GVT's players received earnings that accumulated over time (wealth effects), and changed in accordance to their bets, (3) our shooters always shoot from the same location, whereas GVT's were forced to move after each shot, which enables us to control for the possibility that a player's hit probability varies solely due to shot location,[31] (4) for each shooter we collect 300 shots per session, rather than GVT's 100, which gives our tests substantially more statistical power (see Appendix B.2), (5) we are able to collect multiple sessions of 300 shots from our shooters, across a six month period, in contrast to GVT's single session for each subject. As such, we are able to test not only for the existence of hot hand shooting on the individual level, but also whether hot hand performance can be successfully predicted out of sample.

[30] For example, by forcing a player to predict his own performance before each shot, a separate task he may not be used to, one may divide his attention (Kahneman 1973) or disrupt his flow or rhythm (Csikszentmihalyi 1988), arguably making it less likely that hot hand performance emerges.

[31] Shooters have different probabilities of success at different locations (distance and shot angle). Thus, having players shoot from multiple locations would require one to control for their probability of success at each location. Fixing the shot location increases the power of the design for a given sample size. One might argue that if a player shoots from the same location, he may be able to use the outcome of the previous shot to calibrate the next shot, which could artificially induce serial correlation in a way that would not be reproducible in a game setting (aside from free throw shooting). This concern should be lessened by the fact that, using the empirical strategy we outline in Section 3, we find evidence of a hot hand effect in the original study of GVT, where the shot location changed after each shot. Further, in our design, if a shooter could calibrate, we would expect to see him improve over time, rather than remain at the target 50 percent hit rate for the task, but this is not what we find. In particular, in Section 4.1 we find no evidence of improvement (although we find players shoot significantly worse in their first 3 shots, both in our study and all extant studies).

3 Empirical Strategy

3.1 Hot hand statistics

If the hot hand exists, the underlying mechanism is likely based on (1) a shooter shifting to a relatively higher performance state (due to, for example, internal or external factors leading to increased focus, attention, or motor control [Churchland et al. 2006; Csikszentmihalyi 1988; Kahneman 1973]), (2) a shooter performing better following recent success (due to, for example, positive feedback leading to increased confidence [Bandura 1982]), or some combination of (1) and (2). Because either mechanism involves a player temporarily shooting with a higher probability of success, we should expect hot hand shooting to translate into a tendency for hits to cluster.

We introduce three statistics that not only measure the tendency of a player to produce hit streaks, but also represent the particular patterns typically associated with hot hand shooting, and have improved power over previous statistics (see Appendix B.2 for a discussion of power).[32] Loosely stated, our statistics measure: (1) how often a player is on a "hot" streak, (2) a player's shooting percentage conditional on having a recent streak of success, and (3) the duration of a player's most exceptional hot streak. In addition, we employ a commonly used measure of first order-serial dependence, the number of runs.

Definitions

Because our statistics are defined in terms of streaks, we first define a streak. Let S be the set of shots indices and $\{x_s\}_{s \in S}$ the sequence of shot outcomes, where $x_s = 1$ if shot s is a hit, and $x_s = 0$ if it is a miss. A streak occurs at shot s if a player has just completed k or more shots with the same outcome, i.e. if $x_{s-1} = x_{s-2} = \cdots = x_{s-k}$. For the analysis we report below we set $k = 3$ based on evidence that the intuitive sense of the emergence of a streak generally begins at the third successive event (Carlson and Shu 2007), the fact that previous studies have focused on streaks of at least three hits as indicative of hot hand shooting (Koehler and Conley 2003; Rao 2009b), and for reasons of statistical power. This threshold allows us to expect a player to have more than 30 shots per session in this category, given our design target of 50 percent hits.[33]

[32] For a discussion of the patterns, see Gilovich et al. (1985); Gould (1989); Tversky and Gilovich (1989a); Wardrop (1999). For a summary account, see Reifman (2012).

[33] The results we report also happen to hold if the threshold for a streak is set at four successive hits; setting the threshold higher leads to insufficiently powered tests, because fewer than 9 observations (in 300 shots) are expected

The first conception of the hot hand, that hit streaks occur more often for a player than would be expected if that player had a constant hit rate, is captured by the *hit streak frequency statistic*, H_F, defined as the relative frequency of shots that immediately follow a streak of hits:

$$H_F := \frac{|S_H|}{|S|}$$

where $|\cdot|$ counts the number of shots in a set of shots, and S_H is the subset of shots s that immediately follow a streak of hits, i.e. $S_H := \{s \in S : x_{s-1} = x_{s-2} = x_{s-3} = 1\}$.[34] While the calculation of $|S_H|$ involves overlapping shot windows, it can be shown to have a normal asymptotic distribution (see Appendix B). To illustrate, we calculate that a shooter with a constant 50 percent hit rate is expected to take approximately 37 of 300 shots immediately following a streak of hits, yielding $H_F = .12$.[35]

The second conception of the hot hand, that a player's conditional hit rate immediately following a streak of hits is better than would be expected for a player with a constant hit rate, is captured by the *hit streak momentum statistic*, H_M, defined as the shooting percentage on the shot immediately following a streak of hits (i.e. three or more hits in a row):[36]

$$H_M := \frac{\sum_{s \in S_H} x_s}{|S_H|}$$

Before defining the last two statistics we must set some additional notation. Call a sequence of consecutive hits flanked by misses (or the start or end of the session) a *run* of hits $H_0 \subset S$.[37] Further, let $\mathcal{H}_0 \subset 2^S$ be the collection of all runs of hits, and for $H_0 \in \mathcal{H}_0$, let $|H_0|$ be defined as the length of the run, i.e. the total number of shots in the run of hits. To illustrate, the sequence of shot outcomes 0011010111 has $|S| = 10$ shots, and three runs of hits $\{3, 4\}$, $\{6\}$, and $\{8, 9, 10\}$, with lengths 2, 1, and 3, respectively.

from a 50 percent shooter.

[34] Because it is impossible for the first three shots to immediately follow a hit streak, in the results section we report $H_F := \frac{|S_H|}{|S|-3}$.

[35] Under the assumption that a shooter has a constant hit rate of p, for a finite sample, we can show that $E|S_H^\ell|$, the expected number of shots which immediately follow a sequence of ℓ or more consecutive hits, satisfies $E|S_H^\ell| = \sum_{k=\ell}^{|S|} p^k[(|S| - k)(1 - p) + 1]$.

[36] Here, $|H_M|$ is undefined if $|S_H| = 0$. While H_M is equal to the ratio of two asymptotically normal random variables, in simulations its distribution is symmetric and well-behaved. In particular, $H_M = |S_H^4|/|S_H^3|$ where $S_H^k := \{s \in S : x_{s-1} = x_{s-2} = \cdots = x_{s-k} = 1\}$, and $|S_H^k|$ can be shown to have a normal asymptotic distribution (see Appendix B)

[37] More formally, H_0 is a run of hits if it consists of consecutive indices $s \in S$ where $x_s = 1$ for all $s \in H_0$ and $x_s = 0$ for $s \in \{s_{min} - 1, s_{max} + 1\} \cap H_0^C$, where $s_{min} := \min_{s \in H_0} s$ and $s_{max} := \max_{s \in H_0} s$

The third conception of the hot hand, that hit streaks are longer than would be expected if a player had a constant hit rate, is captured by the *hit streak length statistic*, H_L, defined as the length of the longest run of hits[38]:

$$H_L := \max_{H_0 \in \mathcal{H}_0} |H_0|$$

The distribution of H_L can be approximated using a normal distribution (see Appendix B). To illustrate, the expected maximum run length in a 300 shot session, from a shooter who has hit half of his shots, is $E(H_L) = 7.25$, with H_L exceeding 10 in fewer than 5 percent of sessions.[39]

For each of the hit streak statistics there is a symmetrically defined miss streak statistic. If a player has significantly large hit streak statistics, but miss streak statistics that are not significantly different than expected, this allows us to conclude that a player has the hot hand and not the cold hand (see Section 4).

As we will be testing multiple statistics, we must address the issue of multiple comparisons. Because the hit streak statistics are not independent of one another we do not use standard corrections (such as Bonferroni), which are too conservative.[40] Instead, because each statistic is intended to measure a different dimension of the same underlying construct, we reduce the three hit statistics to a composite statistic H, a linear combination equal to their first principal component, which we compute from the joint distribution of the statistics for a given player's data (as generated by our permutation scheme defined below).[41]

In addition to the hit streak statistics, we consider a statistic that appears prominently in the hot hand literature (Gilovich et al. 1985; Koehler and Conley 2003; Wardrop 1999), the *runs statistic*, R, which counts the total number of runs of hits and misses together:

$$R := |\mathcal{H}_0| + |\mathcal{M}_0|$$

[38] This statistic was suggested by Wardrop (1999), who approximated its distribution numerically and showed that it can be an effective test for the existence of hot hand shooting when the alternative model of the data generating process involves long streaks.

[39] In our statistical tests below we instead numerically approximate the exact distribution of H_L, as the convergence to a symmetric normal distribution is slow.

[40] If a player has a higher hit rate, each statistic will have a higher value. In fact, even for a player with a fixed probability of success, we find that the average pairwise correlation between the statistics is around .5 (as computed from the joint distribution under the null, generated by permutation using the methods outlined further below).

[41] The first eigenvector from the decomposition weights each statistic nearly equally (the hit streak length statistic is weighted a bit less).

where $\mathcal{M}_0 \subset 2^S$ is the collection of all the runs of misses, defined analogously to $\mathcal{H}_0$. That this statistic is a measure of first order serial dependence is illustrated by the fact that the number of runs in a sequence of binary outcomes is simply the number of alternations (pairs of consecutive shots with different outcomes), plus 1.[42] If it is more likely than expected that a hit follows a hit (or that a miss follows a miss), the runs statistic will be *lower than expected.* The expected number of runs from a shooter who has hit half of his shots, in a 300 shot session, is $E(R) = 150.5$, with R falling below 136 in no more than 5 percent of sessions.[43]

3.2 Statistical Test Procedure

Under the null hypothesis (H_0) that a player does not get hot (or cold), the player's shooting performance is a sequence of iid Bernoulli trials with a fixed probability of success. While a player's true success rate is unknown to the experimenter, conditional on the number of successes $N := \sum_{s\in S} x_s$, the shot outcomes are *exchangeable*, i.e. all orderings of the shot outcomes are equally likely under H_0. This means that for a single player's realization of a sequence of 300 shots, an exact (discrete) distribution exists for each statistic outlined above under H_0, by exchangeability; enumerating each permutation of the player's shots, and calculating the value of the test statistic gives this distribution. For a test statistic T defined on the sequence of shot outcomes $\{x_s\}_{s\in S}$, the (one-sided) test with a significance level α is defined by the family of critical values $\{c_{\alpha,k}\}$; if the sequence has $N = k$ successes, $c_{\alpha,k} \in \mathbb{R}$ is the smallest value such that $\mathbb{P}(T \geq c_{\alpha,k}|N = k, H_0) \leq \alpha$, and therefore, $\mathbb{P}(\text{reject}|H_0) = \sum_{k=1}^{|S|} P(T \geq c_{\alpha,k}|\sum_{s\in S} x_s = k, H_0)P(N = k|H_0) \leq \alpha$. While the enumeration required to calculate $\mathbb{P}(T \geq t|N = k, H_0)$ is computationally infeasible, the exact distribution of the test statistic can be numerically approximated to arbitrary precision with a

[42] More precisely, $R := 1 + \sum_{s=1}^{|S|-1}[1 - x_s x_{s+1} - (1 - x_s)(1 - x_{s+1})]$.

[43] Assuming all arrangements equally likely, there is a normal asymptotic distribution for the runs statistic (Mood 1940). The well-known Wald-Wolfowitz runs test uses this normal approximation. We instead use a numerical re-sampling scheme that allows us to approximate the exact p-values of each test statistic to any precision we choose (see Section 3.2). For the number of shots taken in a single session of the current study, the normal approximation of the distribution of the runs statistic is adequate (the p-values it generates are nearly identical to the exact p-values). Nevertheless, when pooling a single player's shooting data from different sessions the normal approximation cannot be used, unless one is willing to assume shooting performance is iid across sessions.

Monte-Carlo permutation of that player's shot outcomes (Ernst 2004; Good 2005).[44,45] We report corresponding one-sided p-values throughout Section 4, as the alternative hypothesis of hot hand shooting establishes a clear ex-ante directional prediction.

Discussion of our Empirical Strategy

The original study of GVT considered three statistical tests: (1) first order serial correlation, ρ, and the number of runs, R—both of which measure the same pattern in shooting, and yield equivalent tests (see Appendix B.1 for a demonstration), (2) a conditional probability test which compares the hit rate immediately following a streak of exactly k hits to the hit rate immediately following a streak of exactly k misses, for $k = 1, 2, 3$, and (3) the variation in the hit rate in four-shot windows.

The statistics we use more tightly correspond to common conceptions of hot hand shooting, substantially increase statistical power, and improve identification: (1) the hit streak length (H_L) and frequency (H_F) statistics can detect the hot hand in a player who occasionally (or frequently) becomes hot and produces extended hit streaks, but who also excessively alternates between hits and misses when not on a hit streak—a combination of patterns that can go undetected by first order serial correlation, and other measures,[46] (2) in testing for improved performance on the shots that immediately follow a streak of hits, the hit streak momentum statistic's (H_M) null distribution is generated by harnessing the entire sequence of shots; this provides the test greater power than GVT's conditional probability test (typically a 10-20 percentage point increase in power; see Appendix B.2), which is expected to use only around 25 percent of the 100 shots from each shooting session,[47] (3) we can identify hot hand shooting, rather than cold hand shooting, as the source of performance variation by using symmetrically defined versions of our statistics for streaks of misses; by contrast, GVT and more recent studies cannot determine whether a statistically

[44] We performed simulations to verify that our code accomplished this (not reported here). For example, for each statistic we determined the 5 percent critical threshold $c_{.05,k}$ by permutation for each k of the 300 possible realized success rates and with this we find that in slightly fewer than 5 percent of our simulated $Bernoulli(p)$ 300-trial experiments with a fixed theoretical success rate p the null was rejected using a test that, in each experiment, selects the permutation test's critical threshold $c_{.05,k}$ corresponding to the player's realized success rate $k/300$ generated by the Bernoulli process for that experiment. This holds for a range of underlying fixed Bernoulli success rates and critical region sizes.

[45] For the test statistic H_M, if the number of successes k is small, then for a fraction of permutations H_M may not be computable as a player may never hit three (or more) shots in a row in a sequence of 300 shots.

[46] Relatedly, another way that a player can have both a hot hand and zero serial correlation in shot outcomes is if hit streaks exhibit persistence (positive serial dependence of a higher order) and miss streaks exhibit reversal (negative serial dependence of a higher order).

[47] GVT's test compares on average 12 shots that immediately follow a streak of three hits with on average 12 shots that immediately follow a streak of three misses.

significant result, should it exist, is due to a hot hand or cold hand shooting, (4) Our empirical strategy can accommodate pooled data, both across sessions and across players, and we increase power by collecting between nine and eighteen times as much data per player as the original GVT study, which has been noted as lacking the requisite power to detect a variety of alternative hot hand hypothesis models (see Appendix B.2 for a discussion of power).[48,49] Further, we use our approach to analyze the data from earlier controlled shooting studies: we perform a pooled analysis of GVT's dataset, yielding tests that have a power advantage over those conducted by GVT on individual shooter level data, and we also test for the existence of a hot hand in a composite dataset consisting of the data from all extant controlled shooting experiments.

Finally, to the extent that the hot hand exists, its relevance for decision making depends largely on the magnitude of its effect. While we can measure standard deviations from the mean of hit streak statistics under the null, the hit streak momentum statistic (H_M) also allows us a direct estimate of magnitude. In particular, H_M measures the shooter's hit rate in a potentially "hot" state (immediately following three or more hits in a row), which can then be compared against the shooter's hit rate in a presumably non-hot state (immediately following any other recent shot history). The difference between these two conditional hit rates provides a conservative estimate of the hot hand effect, in that the true magnitude of the effect, should it exist, will be larger than the estimated effect because not every shot taken immediately following three or more hits in a row occurs while a player is in a hot state (see Appendix B.2; also see Arkes 2013, Green and Zwiebel 2013, and Stone 2012 for a discussion in relation to previous work).[50]

[48]The issue of power limitations in GVT's original study has been pointed out by other authors(Albert 1993; Albert and Williamson 2001; Dorsey-Palmateer and Smith 2004; Hooke 1989; Korb and Stillwell 2003; Miyoshi 2000; Stern and Morris 1993; Stone 2012; Swartz 1990; Wardrop 1999), with only one attempt to address the issue (see Footnote 14).

[49]We can pool all of a player's shooting data, compute the hit streak statistic for each session, standardize it by subtracting the mean and dividing by the standard deviation for that session, then calculate the average of the standardized statistics across sessions, and then generate the distribution of this average by permuting shots within session strata, to assure that the results are not being driven by good day/bad day effects. For pooling across players we average across these player averages, permuting shots within all strata defined by each player and session.

[50]As an example, Stone (2012) has demonstrated that a player can simultaneously have a large serial correlation in shooting ability and a sample serial correlation that is, asymptotically, orders of magnitudes smaller (e.g. .4 and .06, respectively).

4 Results

In Section 4.1 we provide a brief summary of overall performance in our panel of shooters. In Section 4.2 we test if the hot hand effect exists at the individual level, and whether it re-occurs across time. In an analysis of the shooting data from our own study, as well as the data from all extant controlled shooting studies, we find that certain individuals get the "hot hand," and that the effect is systematic and substantial; shooter hit rates increase by between 8 and 40 percentage points immediately following a streak of hits. To put these numbers into perspective, the difference between the best three point shooter and the median three point shooter was 10 percentage points for the 2013-2014 NBA season.

In order to examine whether the hot hand is an average effect among players, in Section 4.3 we extend our analysis to pooled shooting data. Because our experimental paradigm allows us to control for session and day-level variations in performance, our analysis is not vulnerable to the unavoidable endogeneity issues that arise when analyzing in-game shooting data. In our study, we find an average hot hand effect among members of our panel in Phase One of our design, and then again in Phase Two, six months later. Moreover, we find an average hot hand effect in every extant controlled shooting study (including GVT), and again when we pool the shooting data from all studies together.

Finally, in Section 4.4 we find that expert shooters' stated beliefs about which of their teammates have a tendency to get hot (or not), based on previous playing experience, is tightly correlated with the hot hand shooting among these teammates in the controlled shooting task.

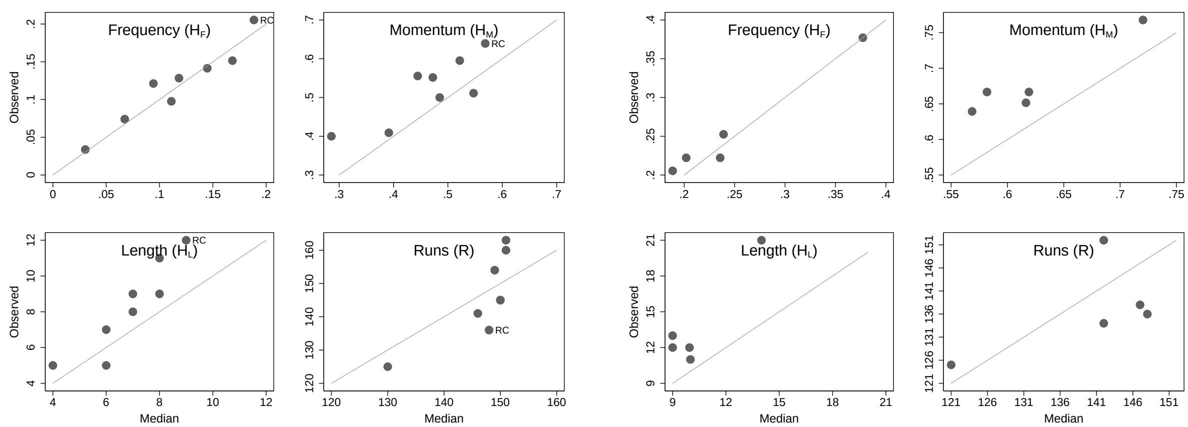
4.1 Overall performance

The average shooting percentage across players in the (balanced) two-phase panel (3 sessions) was 50.08 percent, with a 7.7 percent standard deviation (our design target was 50 percent). By allowing the shooters to warm up before each session, and setting the length of the shooting session to 300 shots, the evidence indicates that players were not subject to warm-up or fatigue effects.[51] In a post-experiment questionnaire the average reported level of fatigue by players was less than 5, on a scale of 1 to 10. Further, the average shooting percentages in the first 150 and second 150 shots

[51] The one minor exception is that for the first three shots there did appear to be a warm-up effect: players shoot significantly worse (37 percent) than the remainder of the session. This is also true in the data of GVT and JNI.

Figure 1: *Scatter diagrams of the three hit streak statistics, and number of runs, with observed value vs. median value under the null distribution for the session.*

(a) *Panel, Phase One (8 shooters, 1 session each)* **(b)** *Shooter "RC" (five sessions)*



were 49.7 and 50.4 percent, respectively, and the average of the session-level differences was not significant. If we instead divide the sessions into three sets of 100 shots, the averages were 50.5, 49.8, and 50.0 percent, respectively, and the average of the session-level differences for each of the three possible comparisons was not significant.[52,53]

4.2 Identification of the hot hand at the individual level

Our Study

Figure 1a reports the hit streak statistics for Phase One of our panel, in which each of eight shooters performed a single 300 shot session. In each cell, each shooter's corresponding hit streak statistic is plotted against its median under the null (based on the number of hits in his 300 shot attempts). One shooter, whom we refer to as *RC*, and whose statistics are labeled in the figure, stood out in Phase One because he had the most extreme hit streak statistics among the shooters, and because his statistics were significant in the single session.[54] Further, in a multi-item questionnaire administered to the team, RC was identified by his teammates as the player with the greatest

[52] A single player in the panel (not the player RC discussed below) performed significantly better in the first half of his sessions (54 percent) than during the second half (46.2 percent), but this same player also performed better on average in his final 100 shots (49 percent) than his middle 100 shots (46 percent).

[53] For player RC, whom we discuss extensively below, across his 5 identical sessions (1500 shots total), his first half and second half performance was not significantly different (61.7 percent vs. 63.3 percent), and his performance in each 100 shot window did not differ significantly either (61.6, 63.4, and 62.6 percent, respectively).

[54] The composite hit streak statistic, H had a p-value of .08 (one-sided).

Table 1: *The average value of the three hit streak statistics and the number of runs for the shooter "RC" from our study, and the shooter "JNI6" from Jagacinski et al. (1979), with (one-sided) p-values in parentheses.*

	Shooter "RC"			Shooter "JNI6"
	session 1	sessions 2-5	All	9 sessions
Hit Streak Frequency (H_F)	.21	.27	.26	.27***
	(.197)	(.335)	(.134)	(.0011)
Hit Streak Momentum (H_M)	.64	.69***	.68***	.71***
	(.105)	(.009)	(.003)	(.0001)
Hit Streak Length (H_L)	12.00	14.25**	13.50**	9.56***
	(.132)	(.037)	(.019)	(.0005)
Total Runs (R)	136.00*	137.25	137.00	26.67**
	(.090)	(.442)	(.168)	(.0262)

Permutation test (50,000 permutations, with session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided)

potential for the hot hand, based on their experience playing with him in games.[55] Thus, when we followed up with the players in our panel six months later, it became possible to use RC's Phase Two shooting data to test if teammate perception of in-game hot hand performance, as well as hot hand performance in the shooting study six months prior, could predict hot hand performance (out of sample). To maximize the identifying power of our test, without informing RC of the purpose, at the end of Phase Two we solicited more sessions from him than from the other players.

Each cell of Figure 1b plots one of RC's hit streak statistics, across each of five sessions, against its median under the null (based on the overall number of hits in that session).[56] RC's hit streak length (H_L) and momentum statistics (H_M) are greater than their respective medians under the null (gray lines) in every session; this would occur in around three out of every hundred studies, for each statistic, if the null hypothesis were true ($p = .031$, one-sided binomial test).[57]

By pooling the shots from all of RC's sessions together, we can perform a more powerful test of

[55] The teammates were not aware of the goal of the study, did not witness RC shoot in the task, and were not informed of his performance.

[56] Due to a time conflict, in one instance RC requested to shoot two of his sessions successively in a single day. While we granted his request, we exclude this session from our analysis because it is conducted under conditions not strictly identical to those of his other sessions. If we instead include this "double session" in the pooled analysis of RC's sessions the significance of his statistics do not change. See Figure 8 in Appendix C for a graph similar to that in Figure 1b, but with this data included (five other shooters in our panel were also granted their request to shoot two successive sessions in a single day, and in these cases we followed the same procedure as with RC).

[57] The hit streak frequency (H_F) and runs (R) statistics are not significant ($p = .50$, one-sided binomial test). As discussed in Section 3.2, the runs statistic is essentially equivalent to serial correlation for binary data. It is not significant for RC because his shot outcomes exhibit persistence after streaks of hits and reversals after streaks of misses.

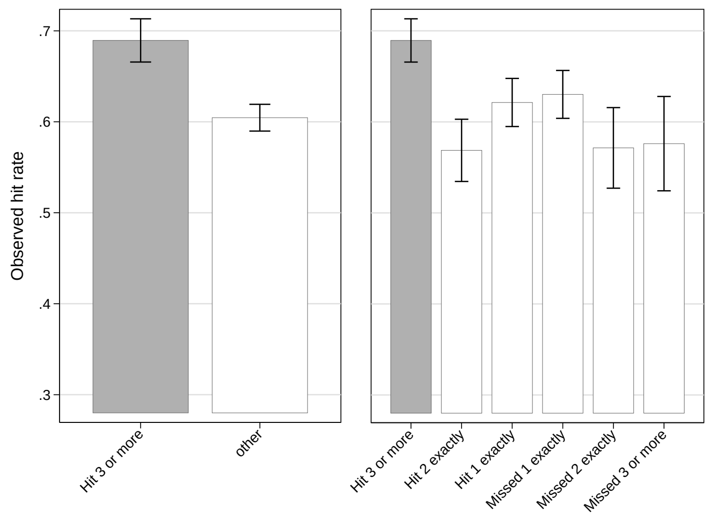


Figure 2: *The shooter RCs hit rate, immediately following 3 or more hits in a row, is higher than after any other recent shot history (with standard errors).*

the hot hand effect. The third column of Table 1 reports each of RC's statistics, averaged across all five of his shooting sessions, with corresponding p-values in parentheses.[58] All of RC's statistics are in the direction predicted by the 'hot hand' hypothesis, and consistent with the results of the binomial test: the hit streak length statistic is significantly larger, by 2.7 shots, than its mean under the null ($H_L = 13.5$, $p = .019$; one-sided permutation test, session strata) and the hit streak momentum statistic is significantly larger, by 6 percentage points, than its mean under the null ($H_M = .68$, $p = .003$), while the frequency and runs statistics are in the predicted direction, but not significant ($p < .20$ for both). When we control for multiple comparisons with the composite hit streak statistic, H, we find it to be highly significant ($p < .01$, one-sided).

While this is the first study to clearly identify the hot hand—and more importantly, at an individual level—an important question is whether the effect size is substantial. Figure 2 presents a bar graph which compares RC's hit rate immediately following a streak of hits with his hit rate immediately following other recent shot histories (with standard error bars). The left panel shows that RC's hit rate increases substantially, by around 9 percentage points, immediately following

[58]The p-value of each statistic comes from a permutation test of the sum of each statistic across sessions (stratified at the session level). Each reported p-value is an approximation of the exact p-value under the exchangeability hypothesis. We do not report Clopper-Pearson binomial confidence intervals for the p-values because with 50,000 permutations, the intervals have a width of less than .001 for most statistics, and a width less than .01 for all.

three or more hits in a row, as compared to any other shooting situation he may find himself in (p=.002, two-sample one-sided test of proportions).[59] Because not every shot taken immediately following three or more hits in a row occurs while a player is in a "hot" state, the hot hand effect will be underestimated. Despite this downward bias, the size of the effect for RC is large enough that it would affect allocation decisions should it occur in a game, and is equal to the difference between the 75^{th} percentile and the top percentile for professional three-point shooters.[60]

Because this effect can be driven by a hot hand or a cold hand (an increased miss percentage immediately following a streak of misses), in the right panel of Figure 2 we categorize shots which do not immediately follow a streak of three or more hits into five mutually exclusive (and exhaustive) recent shot histories: hit the previous two shots but no more, hit the previous shot but no more, missed the previous shot but no more, missed the previous two shots but no more, and missed the previous three or more shots. We see that RC's conditional hit rate immediately following a run of at least three hits exceeds his conditional hit rate immediately following each of the five other recent shot histories (respective p-values for two-sample one-sided test of proportions: .019, .008, .047, .027, and .002). This indicates that the overall contrast observed in the left panel is driven by the hot hand and not the cold hand. We can also test if RC's *miss streak* statistics, which are symmetrically defined to his hit streak statistics, are significant; they are not, corroborating that RC has the hot hand, and not the cold hand. By contrast, statistical tests used in previous studies, including conditional probability tests otherwise similar to our own, cannot separate between the hot hand and the cold hand in this way.

Another confound that could be driving what appears to be a hot hand effect in the results of the conditional probability tests associated with Figure 2 is a selection bias at the session level. In selecting shots that are immediately preceded by a streak of hits, one may over-select shots taken on days when a player is shooting well, thus relatively high observed conditional hit rates may be driven by between-session variation in a player's shooting ability rather than within-session periods of hot hand shooting.[61] To control for this possible confound, in Appendix C.1 we estimate a linear model of RC's hit probability with session fixed effects (which downward bias the estimated effect

[59] The same results hold if one defines a hit streak as beginning at four hits in a row. A benefit of the test reported here is that it includes all of a shooter's data, unlike GVT's conditional probability test.

[60] Data from NBA.com

[61] RC's significant hit streak statistics (Table 1) do not have this issue as they are computed first for each session then averaged across sessions with permutation strata set to the session level.

size). The results of the proportion tests reported in Figure 2 are corroborated, again suggesting that what we observe is a hot hand effect, not a cold hand effect.

The analysis of RC's shooting data demonstrates that an individual can have a substantial hot hand effect that not only systematically re-occurs across time, but can also be correctly predicted—either on the basis of observed performance in previous shooting sessions, or teammate perception of the shooter's in-game performance (see Section 4.4).

A further question of interest is the extent to which there is evidence that other individuals in our panel have the hot hand. Though we have found that the hot hand exists, there is no reason to expect to see hot hand effects from each player in our panel; the subjects were not selected on the basis of shooting ability, or for a reputation of streak shooting, but rather solely on the basis of availability (they were all from the same team). Nevertheless, Figure 1a shows that in Phase One the hit streak length and momentum statistics are each above the median under the null for 7 out of 8 shooters (p=.035 for each, one-sided binomial test). Figure 3a presents a similar plot of the hit streak statistics for each player in the panel, instead with each statistic averaged across the three sessions conducted under identical conditions; the hit streak frequency and momentum statistics, as well as the runs statistic, are on the predicted side of the median for 7 out of 8 shooters (p=.035 for each, one-sided binomial test), while the hit streak length statistic is above median levels for 6 out of 8 shooters (p=.145, one-sided binomial test). In Table 4 of Appendix C we report the percentage of each player's shots that are hits, both immediately following three or more hits in a row, and after all other recent shot histories.

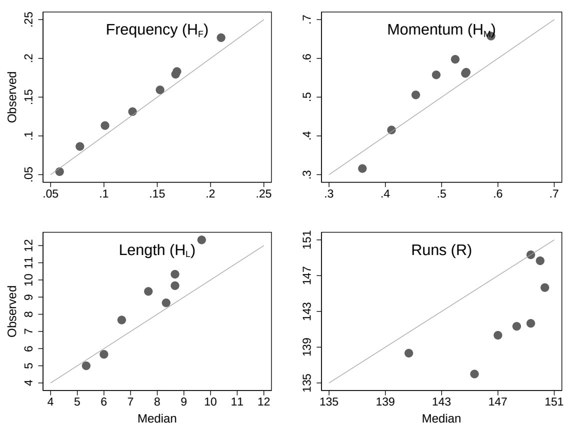
Evidence from earlier controlled shooting studies

With a substantial individual hot hand effect clearly identified in our data, using the statistical measures we have introduced, it is natural to test whether these statistics also detect a similar degree of individual level hot hand shooting in earlier controlled shooting studies. We start with the study which has previously gone uncited within the hot hand literature (Jagacinski et al. 1979; henceforth JNI), as outside of our own study, it provides the richest dataset for individual level testing; each of JNI's six collegiate-level players shot in nine sessions of 60 shots per session.[62]

[62] The subjects shot under three different conditions, "On", "Off", and "Int". We present an analysis of the "On" condition only as the "Off" and "Int" condition involved turning the lights off in the gym after the player shot the ball, and thus were not comparable to other controlled shooting studies (the JNI study was designed to investigate the role of movement and ball trajectory information in the prediction of shooting performance).

Figure 3: *The panel of eight shooters (three sessions per shooter).*

(a) *For each cell, the observed value for the three-session average of the corresponding statistic is plotted against its respective median under the null distribution for the three sessions, for each of the eight shooters in the panel.*



(b) *Average of the hit streak statistics across eight players. The hit streak statistics for each shooter are first normalized in each session to a z-score using the null distribution for that session (one-sided p-values in parentheses).*

	Overall	Phase 1	Phase 2
Hit Streak Frequency (H_F)	.51***	.28	.63***
	(.006)	(.215)	(.006)
Hit Streak Momentum (H_M)	.48***	.66**	.39*
	(.008)	(.028)	(.057)
Hit Streak Length (H_L)	.43**	.54*	.38*
	(.021)	(.069)	(.071)
Total Runs (R)	−.58***	−.10	−.83***
	(.002)	(.387)	(.000)

50,000 Permutations (session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided)

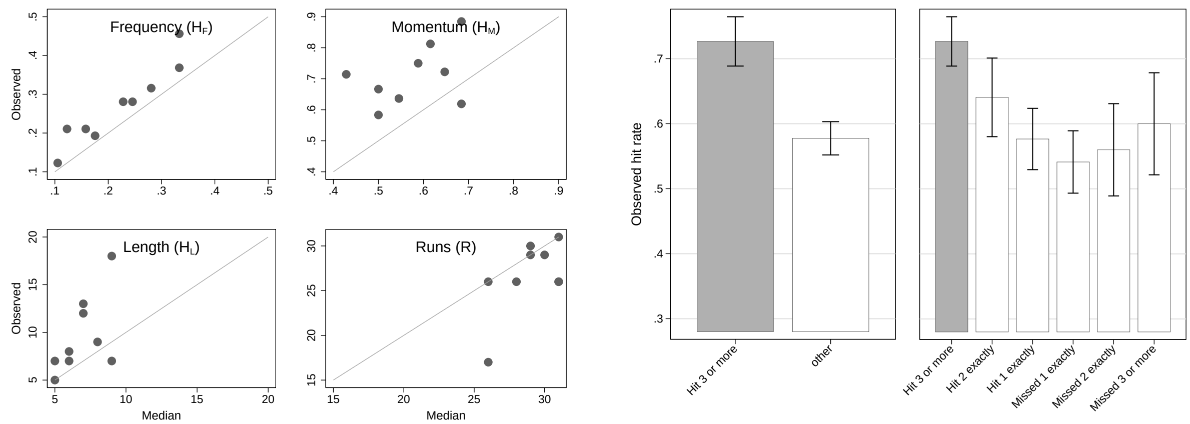
We focus here on one of JNI's subjects, *JNI6*, because his hot hand performance is substantial, and systematically re-occurs in almost every one of his nine shooting sessions. Our analysis of JNI6's shooting data is identical to that performed on RC's. Figure 4a reports that in all nine of JNI6's sessions, the hit streak frequency statistic is above median levels (p=.002, one-sided binomial test), in eight of nine sessions the hit streak momentum statistic is above median levels (p=.02, one-sided binomial test), in seven of nine sessions the hit streak length statistic is above median levels (p=.09, one-sided binomial test), while the runs statistic displays no particular (directional) pattern (p=.5, one-sided binomial test).

As in the analysis of RC, by pooling the shots from all of JNI6's sessions we can perform a more powerful individual-level test of the hot hand effect. For each hit streak statistic we average its value across all nine of JNI6's sessions to provide a more representative test in which the magnitude of the effect in each session is weighted equally. The last column of Table 1, further above, reports that each of JNI6's average hit streak statistics is highly significant (p-values in parenthesis), which constitutes clear evidence that JNI6 has the hot hand, as his hit streak statistics all remain highly significant even after a Bonferroni correction for JNI6 being only one of six players. To get a feel for the magnitude of each hit streak statistic, one can compare it against its median value under

Figure 4: *Evidence of the hot hand for shooter "JNI6" of Jagacinski et al. (1979).*

(a) *In each cell, the observed value for the corresponding statistic is plotted against its respective median under the null distribution for that session, for each of the nine sessions.*

(b) *The hit rate for shooter JNI6 is higher after having hit 3 or more in a row, than after any other recent shot history (with standard errors).*



the null: for frequency (.27 vs. .22), momentum (.71 vs. .56), and length (9.6 vs. 7.0).

As in the case of RC, another way of checking whether JNI6's hot hand effect is not only highly significant, but also substantial, is by comparing his conditional hit rate immediately following a sequence of at least three hits, with his conditional hit rate immediately following any other recent shot history. The left panel of Figure 4b shows that JNI6's conditional hit rate increases substantially, by around 15 percentage points, when immediately following three or more hits in a row, as compared to his conditional hit rate following all other recent shot histories (p=.001, two-sample one-sided test of proportions).

For this test, we can also check whether JNI6's substantial changes in performance after streaks is being driven by the cold hand. The right panel of Figure 4b confirms that JNI6's performance differences are indeed being driven by the hot hand, and not by the cold hand, as his conditional hit rate immediately following a run of three or more hits exceeds his conditional hit rate immediately following each of the five other categories of recent shot histories.[63,64] We can also perform an additional check in which we test JNI6's miss streak statistics. We find that they are not significant,

[63] The corresponding p-values for the two-sample one-sided test of proportions are .107, .006, .001, .015, and .062, respectively.

[64] These effects are not being driven by between-session variation in performance. As with the player RC, to control for between session variation in performance we estimate a fixed-effects linear probability model of JNI6's hit rate, which corroborates these results (see Appendix C, Table 3).

which further corroborates that JNI6 has the hot hand, and not the cold hand.

JNI6 is not the only player from the JNI study with significant hot hand effects. A second player has significant hit streak frequency and runs statistics (.05 level for each, one-sided test).[65] The probability of 2 out of 6 players crossing this significance threshold is .03 (one-sided binomial test).

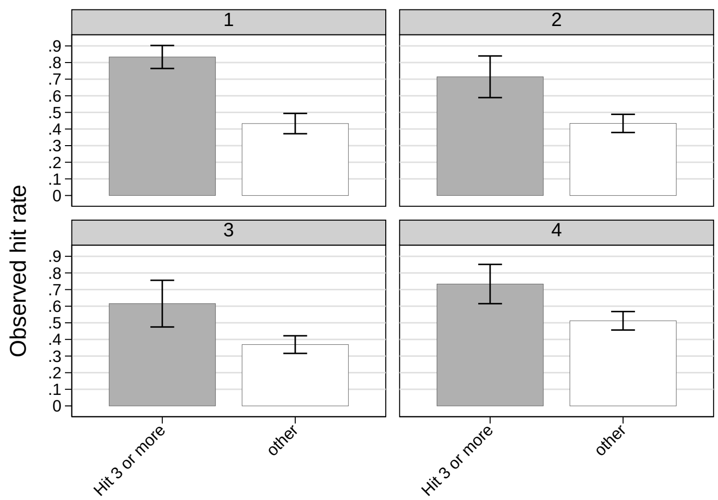


Figure 5: *Four shooters in the GVT experiment have a significantly higher performance after a streak of 3 or more hits in a row, than any other shooting situation.*

In Section 3 we mentioned that with only a single session of 100 shots per shooter, GVT's individual-level statistical tests for the hot hand effect were severely underpowered (see Appendix B.2 for details). Nevertheless, our tests reveal evidence of the hot hand effect in their data, with a magnitude that would be substantial if it were to maintain in a larger sample. In particular, the conditional hit rate of 8 out of their 26 shooters increases by at least 10 percentage points immediately following a sequence of three or more hits. In Figure 5 we present the performance for four of these shooters, whose percentage point increases were 40, 28, 25 and 22 respectively—with the increases significant for all four of these shooters (.05 level, one-sided two-sample proportion test).[66] The probability of this occurrence under the null is $p = .039$ (one-sided binomial test).

[65]The hit streak momentum and length statistics for this player are not significant, though in the top quintile under the null.

[66]For the player with a 40 percentage point increase in shooting percentage, 30 of his shots were taken immediately following a run of three or more hits. For the other players the number of shots taken with this recent shot history were 15, 13, and 14, respectively.

Moreover, the runs statistic for 5 out of 26 exceeds the one sided .05 significance threshold—an event that occurs with a binomial probability of $p = .009$ under the null.[67]

To put these individual-level results into perspective, the validity of GVT's conclusion was temporarily called into question by a study which claimed that a single player, Vinnie "the Microwave" Johnson, had the hot hand while playing for the Detroit Pistons in the 1987-1988 season (Larkey, Smith, and Kadane 1989), until the study was later called into question due to data coding errors (Tversky and Gilovich 1989b). Unlike in the case of Vinnie Johnson, whose shooting data was selected from a population of more than 300 players precisely because he was widely believed to be one of the hottest shooters in the NBA, we find clear evidence of sizeable hot hand shooting among players in a much smaller group, who were selected only on the basis of availability.

4.3 Pooled Analysis: Average Effects

The individual-level analysis reported in Section 4.2 not only allows for a test of the existence of the hot hand in individuals, but also provides evidence of the heterogeneity of the hot hand effect across individuals. While we have demonstrated that some players systematically get the hot hand, other players appear to always shoot with the same ability, and still others actually under-perform immediately following a streak of hits. For these reasons a pooled test can only provide limited information about the existence of hot hand shooting in individuals; if one observes a pooled hot hand effect, then this suggests that at least one individual in the sample has the hot hand, whereas if no pooled effect is observed, without further information, one cannot know whether or not there are individuals with the hot hand in the sample. On the other hand, a pooled analysis *can* answer the question of whether the hot hand is an average effect of the shooters in a sample. We find clear evidence of an average hot hand effect, across all extant controlled shooting studies. This result is surprising because it indicates that the hot hand is not only a property of a few rare players, but also a property of the average shooter.

Table 2 reports the across-player average of the hit streak statistics (in standardized units), for each controlled shooting study, with the shooters from our panel in column 1, GVT's shooters in column 2, JNI's shooters in Column 3, and all shooters together in column 4. Each hit streak

[67] A further reason why individual-level tests are underpowered in the GVT data is that many of their players had a low overall hit rate (less than 35 percent), which reduced the amount of testable data for these players. There is some evidence of the cold hand in these players, e.g. one player never made more than two consecutive shots.

Table 2: *The average value of the (normalized) hit streak statistics across players in our study, Gilovich et al. (1985), and Jagacinski et al. (1979), with (one-sided) p-values in parentheses. The hit streak statistics are constructed by first normalizing the statistic in each session to a z-score using the null distribution for that session, then averaging across sessions for each player to create an average z-score for each player, then averaging across players.*

	Panel	GVT	JNI	Pooled †
Hit Streak Frequency (H_F)	$.51^{***}$	$.42^{**}$	$.24^{**}$	$.29^{**}$
	(.006)	(.017)	(.038)	(.013)
Hit Streak Momentum (H_M)	$.48^{***}$	$.37^{**}$	.06	$.24^{**}$
	(.008)	(.031)	(.341)	(.027)
Hit Streak Length (H_L)	$.43^{**}$	.24	.13	.16
	(.021)	(.109)	(.164)	(.112)
Total Runs (R)	$-.58^{***}$	$-.21$	$-.37^{***}$	$-.21^{**}$
	(.002)	(.144)	(.003)	(.046)

Permutation test (50,000 permutations, with session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided)

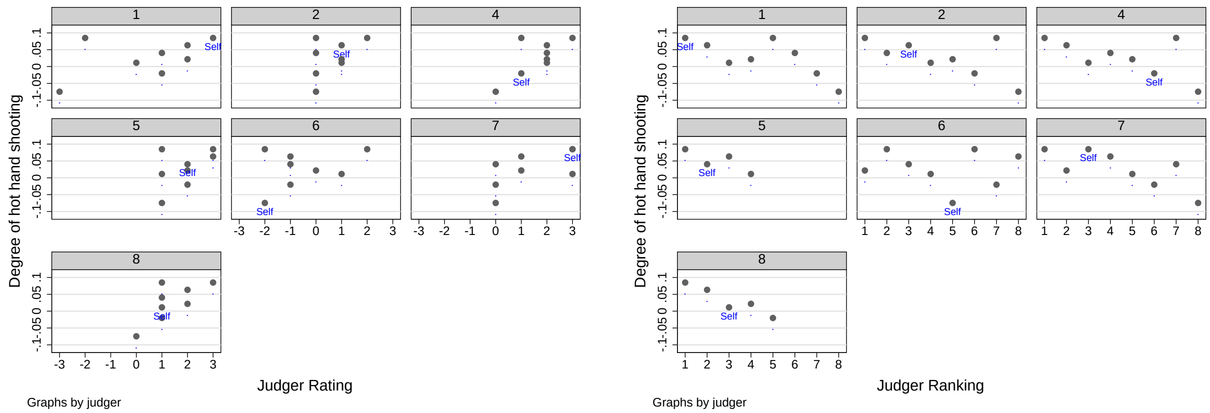
†The panel has 3 sessions of 300 shots per shooter (8 shooters). GVT has 1 session of 100 shots per shooter (26 shooters). JNI has 9 sessions of 60 shots per shooter (6 shooters). The pooled analysis of all studies includes ad-hoc 300-shot sessions that were conduced with players who did not participate in the two-phase panel (5 shooters).

statistic in the pooled data is generated by first standardizing the session level statistic for each player by subtracting the mean and dividing by the standard deviation for that session (under the null distribution generated for that session), then averaging across sessions for each player, and finally averaging across players. The null distribution is generated under the assumption of exchangeability within sessions (but not across, to avoid good day/bad day effects).

When considering just the players in our panel we find a highly significant hot hand effect. All three of the hit streak statistics, as well as the runs statistic, are highly significant. Moreover, this effect is not driven by any single session; to the contrary, Figure 3b shows that hot hand performance in Phase One predicted the presence of hot hand performance in Phase Two (out of sample), six months later. In the pooled analysis of GVT's shooters, we observe significant hit streak frequency and momentum statistics, and JNI's shooters exhibit significant hit streak frequency and the runs statistics. Finally, we pool together all available shooting sessions (including ad-hoc sessions conducted with players not included in our panel), revealing an average hot hand effect among players, albeit with a modest .2 to .3 standard deviation increase in performance, which is comparable to the average effect size detected in game data (Arkes 2010; Bocskocsky et al.

Figure 6: *Judger ratings and rankings predict which players have the tendency to exhibit a hot hand.*

(a) *Hot hand performance vs. judger ratings* **(b)** *Hot hand performance vs. judger rankings.*



2014; Yaari and Eisenmann 2011). This modest average effect size belies the substantial individual effect size observed in particular shooters.

4.4 Expert Beliefs

At the completion of Phase Two of the experiment a detailed questionnaire was administered to the shooters, in which they were asked to asses the degree of hot hand shooting in their teammates (who were also shooters in our experiment) in multiple ways. In general, we find that these expert players can identify which of their teammates have a greater tendency to exhibit hot hand shooting. Because a shooter's session was always observed by himself (and in some instances also by one other member of the panel), this experience could have influenced his stated beliefs about the degree of his own hot hand shooting relative to that of others. Thus, in order to test beliefs based only on previous playing experience, and not on observation of shooting performance in our task, we also repeat our analysis, but excluding a subject's beliefs about a shooter if he observed any of that shooter's sessions (see footnotes below).[68] This analysis yields similar results, indicating that subjects' beliefs, based only on previous playing experience, can predict performance in our controlled shooting task, out of sample.

[68] Despite subjects not being aware of the purpose of the study, as a precautionary measure, before each shooting session it was stressed in the instructions not to communicate any information about the shooting session to others. Subjects were never provided hot hand performance information prior to the survey.

Subjects were first asked to create a rating of how much more or less likely each of the shooters in the panel is to make the next shot following three or more made shots in a row (relative to how each usually shoots) on a scale of -3 (much less likely) to +3 (much more likely). Each cell in Figure 6a contains a plot of the hot hand performance of shooters against the ratings of a particular subject (henceforth "judger"). The vertical axis corresponds to the actual degree of hot hand shooting performance in each of the judger's teammates, defined as the difference between that teammate's conditional hit rate immediately following three or more hits in a row, and his overall hit rate. The horizontal axis corresponds to the judger's respective rating. Judgers' responses indicate that they all believe that at least one other member of the panel has a hot hand, but only one judger believes that all eight shooters do. On average, they rate other shooters as slightly more likely to hit the next shot immediately following three or more hits (+ 1.0), and the shooter RC is rated as much more likely (+ 2.8). Further, six of the seven shooters that judgers' rate (on average) as having a hot hand, do directionally exhibit a hot hand in performance. Likewise, the one shooter that judgers rate as having the opposite of a hot hand (an "anti-hot" hand) does directionally exhibit an anti-hot hand.[69] In addition, judgers' ratings of players are highly correlated with the players' actual degrees of hot hand shooting performance in the shooting task. In particular, eight out of eight judgers' correlations are positive—an event which has a binomial probability of .004 under the assumption that judger rankings are made at random. Further, the average judger correlation of 0.49 is found to be highly significant ($p < .0001$) in a permutation test conducted under the same null, and stratified at the judger level.[70,71]

Judgers were later asked to create a ranking of all of the shooters in the panel in terms of whose shooting percentage rises the most immediately after having made at least three shots in

[69] One can conduct a test of whether the average ratings of shooters (across judgers), correctly predict whether each shooter's performance directionally exhibits the hot hand or the opposite (the "anti-hot" hand). The sign of the hot hand effect is correctly predicted for seven out of eight shooters, which is significant under the null hypothesis that the sign of the average rating is determined at random (p=.035, one-sided binomial test). When we exclude judger-observed ratings, the sign is correctly predicted for six out of eight shooters (p = 0.144, one-sided binomial test), as one of the seven directionally hot hand shooters is believed by judgers to be anti-hot.

[70] When we exclude judger-observed sessions, the average rating for all shooters is +0.8 (for RC it remains +2.8). Further, six out of seven subjects' correlations are positive (p=.064, one-sided binomial test), with a highly significant average correlation of .38 ($p = .016$, stratified permutation test).

[71] Here it is assumed that judgers uniformly randomize among the ratings that we observed them make at least once. If we instead assume that they randomize across all possible ratings the statistical significance is even stronger.

Figure 7: *Scattered display of the eight shooter names used for the ranking task (real names used in actual questionnaire).*

a row.[72,73] The names of the players were presented visually in a graphically scattered fashion so as to minimize the possibility of priming effects (see Figure 7). Figure 6b is analogous to 6a, but with the degrees of hot hand performance across shooters plotted against each judger's hot hand rankings, rather than ratings. Consistent with responses in the ratings questions, judgers' by far rank RC as the hottest shooter (1.1 average rank vs. 3.4 for next highest average rank), and overall rankings are highly correlated (inversely) with the differences we observe between the conditional hit rates of each shooter immediately following three or more hits, and the shooter's overall hit rate in our experiment. In particular, seven out of seven judgers' correlations are negative (p = .008, one-sided binomial test), and the average correlation is $-.60$ ($p < .0001$, stratified permutation test).[74]

Overall, we find that the beliefs of our expert players about the hot hand shooting of their teammates, which were formed during *prior* playing experience, correctly predict which of their

[72] After the rating task, but before the ranking task, judgers were additionally given a slightly more involved series of questions in which they were told each shooter's base rate (overall hit rate), and asked to provide a numerical percentage estimate of the shooter's conditional hit rate immediately following three or more hits in a row. Interestingly, the implied difference in hit rates is not correlated with the ranking task, or the rating task, which are both highly correlated with each other.

[73] One judger stopped filling in the questionnaire before the ranking questions (after the rating questions).

[74] Excluding judger-observed sessions, the judgers' average ranking of RC is 1.2, with the next highest average rank being 2.8. In additional, six out of six judgers' correlations are negative ($p = .016$, one-sided binomial test) and the average correlation is $-.60$ ($p < .01$, stratified permutation test)

teammates have a greater tendency to exhibit the hot hand in our controlled shooting task. Accordingly, it appears that players who have a tendency to get (more) hot in controlled shooting environments, may also have a tendency to get (more) hot in games.

5 Discussion

The original demonstration of the hot hand fallacy by Gilovich, Vallone, and Tversky (1985) has long been considered the strongest supporting evidence of the fallacy. Nevertheless, we have used the best available data to demonstrate that the belief in the hot hand is not a fallacy, and that the original evidence is actually in line with this assessment. In particular, our analysis reveals significant hot hand effects in all extant controlled shooting, and these effects are substantial for some players. In Section 5.1 we discuss what our results, in conjunction with those of in-game shooting studies, indicate about existence of the hot hand effect in basketball games. These results also raise questions about the relevance of the hot hand fallacy in decision making more generally, and in particular, its robustness to expertise, incentives, and learning. In Section 5.2 we discuss why the existing literature related to the hot hand fallacy cannot fully answer these questions. Finally, in Section 5.3 we briefly review evidence in support of the hot hand effect in other domains of skilled performance.

5.1 Hot hand effect in games

A number of recent studies have found evidence of a possible hot hand effect in pooled analyses of professional in-game basketball shooting data, which in theory does speak to the hot hand fallacy, because we know that expert practitioners in this domain believe in hot hand shooting, and act accordingly. Arkes (2010) found evidence consistent with hot hand shooting in free-throw (dead ball) data, observing that a player's success rate is higher following a hit than a miss. Yaari and Eisenmann (2011) then replicated this result with a slightly different analysis, and more data. Bocskocsky et al. (2014) use field goal (live ball) data, with the most extensive list of controls so far possible, and find that, in aggregate, shooting performance (net of expectations) is positively correlated with shooting performance in the previous four shots (net of expectations). These studies are consistent with the possibility of substantial hot hand shooting, and in direct contrast with the original results of GVT. However, in line with concerns that some of these authors themselves point out, and as we argue in Section 2.1, any analysis of in-game shooting, whether it be from the free throw line (dead ball) or field (live ball), will be sufficiently vulnerable to omitted variable bias, so as to make it difficult to conclude that observed performance clustering is in fact attributable to hot hand shooting.[75]

To the extent that hot hand shooting cannot be identified, or dismissed, with game data, but only with controlled shooting data, is at first glance disappointing because what we are primarily interested in is the possibility that, in professional basketball games, the hot hand exists and decision makers are able to detect it and respond to it, as they claim. While we believe we have no choice but to test for the hot hand effect in a sufficiently controlled shooting domain, fortunately, the evidence we present suggests that the hot hand effect likely exists in games and is detectable by decision makers, which is consistent with the results of the in-game studies just summarized. For one, that we find the hot hand effect to be robust across all extant controlled shooting designs, which are all different, suggests that it would generalize to other settings that involve the same physical act. Second, our participants' stated beliefs about which of their teammates have a tendency to become hot, based solely on previous playing experience, is shown in Section 4.4 to be tightly correlated with their teammates' tendency to exhibit hot hand shooting in our experiment. That players' perceptions of hot hand shooting of teammates based on previous experience can be used to accurately predict hot hand shooting in our controlled experiment suggests that shooters who tend to be (more) hot in games also tend to be (more) hot in our controlled shooting task, and visa versa. Finally, in a game environment, coaches and teammates, who know each other well, have access to considerably more information than whether a "hot" player's previous shots simply hit or missed; in conjunction with this, they observe how cleanly the ball entered the rim (or how badly it missed), the player's shooting mechanics, ball trajectory, body language, and other signals of the player's underlying mental and physical state, which can help them distinguish between a

[75] It has been pointed out that a direct effect of the hot hand on shooting performance in games may not be detectable due to strategic adjustments by the defense which lead to few shooting opportunities—but indirect effects of the hot hand may be detectable in improved team performance (Dixit and Nalebuff 1991, pg. 7). Aharoni and Sarig (2011) have found that team performance improves when a player has superior recent performance, suggesting that while the defense is able to adjust and prevent an individual shooter from exploiting his hot hand, this adjustment leaves the defense more vulnerable. It is possible that the defense is exhibiting an erroneous hot hand belief and adjusting disadvantageously to the behavior of the offense, but because it has been shown that both offenses and defenses adjust to "hot" shooters (Aharoni and Sarig 2011; Attali 2013; Bocskocsky et al. 2014; Cao 2011; Rao 2009a), if one assumes that these adjustments cancel, this evidence suggests a presence of a hot hand effect in games.

lucky streak of hits, and a hot hand.

5.2 Assessing the Hot Hand Fallacy

A large literature exists investigating the impact of the hot hand fallacy on economic decision making. While these studies typically involve the provision of a unique data set, a creative identification strategy, or an inventive experimental design, in our assessment, they do not provide evidence of the fallacy that is comparable in strength to the consensus view of the evidence provided in GVT's original study, and the studies which followed. GVT not only identified an environment in which one could conclude that the fallacy existed among experts acting in their domain of expertise, with large financial incentives, feedback, and opportunities to learn, they also identified a clear conflict between beliefs and observed performance. Further, the studies involving basketball data which followed GVT demonstrated that this instantiation of the hot hand fallacy is unique in its generalizability, as it appeared to be both massive and widespread (Kahneman 2011).[76] In studies which investigate the impact of the hot hand fallacy in other domains, when the presence of the hot hand fallacy is clearly identified, there is little evidence that the belief is held strongly enough to have a costly impact on decision making; when the impact on decision making appears costly, ability to identify the role of the fallacy is limited.

With the evidence in the crucial basketball domain now reversed, while it is still possible that the hot hand fallacy is massive and widespread in other domains, in our view, the current evidence supporting this notion is lacking. If the absence of evidence is taken as evidence of absence, the scope for the hot hand fallacy to have a meaningful influence on economic (incentivized) decision making in other domains is also limited. Nevertheless, even if the hot hand fallacy is not as strongly held as once believed, this does not mean the fallacy is a cognitive illusion without potential welfare implications. In particular, in many important domains decision makers and other stakeholders may sometimes not have incentive to have accurate beliefs, e.g. the news coverage, political pressure, and litigation which follows the emergence of cancer clusters that happen to be unexceptional from the perspective of an epidemiologist (Thaler and Sunstein 2008).

[76] More recent evidence that coaches and players make decisions based on hot hand beliefs can be found in Aharoni and Sarig (2011); Attali (2013); Bocskocsky et al. (2014); Cao (2011); Rao (2009a). With respect to the semi-professional players in our study, their responses to a post-experiment questionnaire reveal that they too believe in the hot hand (see Section 4.4).

Field evidence of the hot hand fallacy

In field settings it is difficult to provide as strong of evidence for or against the hot hand fallacy as was seemingly provided by GVT. To do so one must first pin-down the patterns in the outcomes of the data generating process that decision makers have beliefs about, then identify the beliefs that decision makers have about these patterns, and finally, test whether observed patterns and beliefs are in conflict. Moreover, to have evidence on par with that which existed in the basketball domain, one would want to show that these erroneous beliefs are strongly held, by, for example, demonstrating that they are costly to hold, influence the decision making of experts, or are robust to expert feedback and opportunities to learn. These desiderata have not been simultaneously met by any extant field study relating to the hot hand fallacy.

A large body of field evidence in the domains of casino gambling (Croson and Sundali 2005; Keren and Wagenaar 1985; Sundali and Croson 2006), lottery play (Galbo-Jørgensen et al. 2013; Guryan and Kearney 2008), sports wagering (Arkes 2011; Avery and Chevalier 1999; Brown and Sauer 1993; Camerer 1989; Durham et al. 2005), and financial markets (Barberis and Thaler 2003; De Bondt 1993; De Long et al. 1991; Kahneman and Riepe 1998; Loh and Warachka 2012; Malkiel 2011; Rabin and Vayanos 2010) identify anomalies that have been attributed to the hot hand fallacy. In this discussion we focus on the domains of casino betting, lottery play, and sports betting, which are each amenable to a test of the hot hand fallacy because, unlike the literature in finance, the scope for alternative explanations is more limited, as choices are relatively simple, terminal payoffs are realized in a finite and proximate horizon, and there is no feedback of individual decisions into the behavior of the data generating process (Sauer 1998; Thaler and Ziemba 1988).[77] While it

[77] In the finance literature anomalies such as the fund flow puzzle, as well as momentum and reversals in asset returns, have sometimes been attributed to the hot hand beliefs of investors. It is posited that these investors incorrectly extrapolate past returns, or over-infer from recent returns that there has been a shift in regime (Barberis et al. 1998; DellaVigna 2009; Mullainathan 2002; Rabin 2002; Rabin and Vayanos 2010). However, the mechanisms behind these anomalies are difficult to isolate given the relative richness of financial market settings (Brav, Heaton, and Rosenberg 2004; Brav and Heaton 2002).

The *fund-flow puzzle* (Chevalier and Ellison 1997; Sirri and Tufano 1998), in which investors prefer mutual funds with superior recent returns, has been attributed to a hot hand belief by investors (Rabin and Vayanos 2010). This phenomenon can also be explained by limited attention of investors and the salience of top performers (Barber and Odean 2008; Bordalo, Gennaioli, and Shleifer 2012). Moreover, there are also rational accounts based on investors attempting to identify skillful managers have been proposed (Berk and Green 2004), and there is evidence that fund manager skill exists (Chevalier and Ellison 1999) as well as *performance persistence* for highly skilled managers (Jagannathan et al. 2010), although the evidence is mixed (Carhart 1997; Hendricks et al. 1993). More recently, some predictions of the Berk and Green model have been disconfirmed (Fama and French 2010), and it cannot explain, overall, why retail investors are investing in active funds to begin with (French 2008).

Models of investor behavior in which agents have hot hand-like beliefs and over-extrapolate recent price trends may be argued that participants in these settings are not representative of the general population, or that an individual's behavior in these settings is recreational and not representative of his or her behavior in other settings, it has been shown that economic incentives do influence behavior in these settings (Kearney 2005) and that economic theory can characterize aggregate behavior (Sauer 1998).

In the domain of casino gambling it is reasonable to assume that there is no predictability in the underlying process, and thus a gambler's belief that a pattern of recent success signals future success ('hot hand') is, by definition, a logical fallacy.[78] In surveys, recreational gamblers have been found to believe in the hot hand (aka "luck"), reporting that they will bet more when they feel they are on a hot streak (Alter and Oppenheimer 2006; Keren and Wagenaar 1985).[79] Nevertheless, it cannot be inferred from survey data the degree to which these beliefs influence behavior. Moreover, the meaning of these beliefs is unclear in games that involve a degree of skill, which players may be uncertain about, and may fluctuate (e.g. blackjack). An actual casino gambling setting appears to be a more fruitful source of data. Unfortunately, existing studies of gambling behavior are

(Barberis et al. 1998; De Long, Shleifer, Summers, and Waldmann 1990; De Long et al. 1991; Lakonishok, Shleifer, and Vishny 1994) have been proposed as explanations for the robust evidence of *momentum effects* in stock returns (Fama and French 2012; Jegadeesh and Titman 1993), although the evidence for investor over-extrapolation is mixed (Loh and Warachka 2012). Alternative explanations for the momentum anomaly have been proposed (Lewellen 2002), and a rational account of individual behavior based on learning in an uncertain environment has been shown to be difficult to distinguish from the behavioral account in financial market data (Brav and Heaton 2002). For other perspectives see Fama (1998), LeRoy (2004), and Pástor and Veronesi (2009).

Retail investors, whose behavior can move markets (Barber, Odean, and Zhu 2009a), have been observed to trade too much (Odean 1999), and the possibility that they have a tendency to employ momentum strategies or follow price trends—presumably based on hot hand-like beliefs—has been suggested as an explanation (Barber, Odean, and Zhu 2009b; Odean 1999), though alternative (and complementary) mechanisms of influence have also been proposed (Barber and Odean 2008; Barber et al. 2009b). Individual-level evidence of extrapolation bias does exist in the form of surveys of investors (De Bondt 1998, 1993) and home buyers (Case and Shiller 1988; Case, Shiller, and Thompson 2012), but the mechanism which drives these beliefs is unclear, and the connection to behavior must be inferred.

Patterns in employee over-investment in company stock (in their 401k accounts) has been attributed to over-extrapolation of past performance (Benartzi 2001), though one cannot rule out the role of peer effects (Benartzi and Thaler 2007), or regret driven by social comparison similar to that found in postal code lotteries (Kuhn, Kooreman, Soetevent, and Kapteyn 2011).

[78] Nevertheless, it is perhaps not entirely unreasonable to believe that predictable patterns may exist in casino game mechanisms. Roulette wheels, for example, are physical devices operated by a human hand, there is a conflict of interest between gamblers and the casino, and casinos themselves switch roulette wheels after observing even relatively modest clustering of a single color in recent outcomes (Bar-Hillel and Wagenaar 1991). If a gambler believes the wheel is entirely random (i.i.d.), but does not bet accordingly, so long as the inconsistency is costly, it seems correct to call this a fallacy, but if the gambler has priors that the wheel may not be random, then determining if her beliefs about a particular streak of outcomes is erroneous depends on an assessment of the reasonableness of her priors about the process (e.g. distrust in the gaming commission), and the consistency of her priors with her posteriors, based on what she has observed.

[79] The connection between the belief in luck and the belief in the hot hand is discussed in Falk and Konold (1997) and Ayton and Fischer (2004).

not designed specifically for detecting the hot hand fallacy. The betting changes that have been observed in response to "hot" streaks of success cannot guarantee if a player actually bets more when they are "hot" (i.e. that there are expected costs), and even if one is willing to assume that they do, the betting amount might increase simply because bets are positively correlated with the amount of money available to bet, independent of recent success (Croson and Sundali 2005; Sundali and Croson 2006).[80] More recent evidence indicates that gamblers do not exhibit the hot hand fallacy in their betting patterns following a recent stroke of "luck": Smith et al. (2009) find that poker players bet more after losing than after winning, Narayanan and Manchanda (2012) find that slot machine players bet as if they follow the gambler's fallacy for their own performance rather than the hot hand fallacy, and Xu and Harvey (2014) find that online sports bettors migrate towards bets with more favorable odds after winning, and towards bets with longer odds after losing.

In the domain of lottery play, as in casino gambling, a lack of predictability in the underlying process can also be safely assumed, and therefore only the beliefs and behavior of players need be investigated. In lottery drawing it has been found that (1) relatively more money is placed on a number under a parimutuel pay-out scheme (implying lower expected payout) the more frequently that number has been drawn recently (Galbo-Jørgensen et al. 2013),[81] and (2) "lucky" stores sell more lottery tickets when they have recently sold a winning ticket, implying that players erroneously believe these tickets have a higher chance of winning (Guryan and Kearney 2008). The popularity of hot numbers in parimutuel games is solid evidence that hot hand fallacy is a cognitive illusion that operates in field settings, but the evidence is not sufficient to conclude that the belief is strongly held for two reasons: (1) the game is simultaneous move, so conditional on winning, the cost of holding the belief is unknown, and (2) the chance of winning is close to one in one million. The lucky store effect also demonstrates a mistaken belief in the hot hand, but as in the parimutuel setting, the strength of the beliefs cannot be determined: sales increases are driven by a combination of a substitution away from non-winning stores (unknown cost), and a general increase in sales, which

[80] This correlation may be driven by change in risk preferences arising from a house money effect (Thaler and Johnson 1990), a wealth effect (Ball 2012), or diminished sensitivity to rewards arising from adaptive coding (Padoa-Schioppa and Rustichini 2014) and relative judgement (Stewart, Reimers, and Harris 2014).

[81] Earlier investigations on the gambler's fallacy have been conducted using lottery data. In an analysis of pick 3 and pick 4 numbers games, Terrell (1994) tests for the effect of the gambler's fallacy in a parimutuel setting, while Clotfelter and Cook (1991) and Clotfelter and Cook (1993) test for the effect in a fixed-prize setting. The hot hand fallacy cannot be tested in this setting because the probability of a streak of the same number is vanishingly small.

could arise from increased purchases by existing players (costly) or a media and advertising effect that catches the attention of new customers (costly, but not related to hot hand beliefs).[82]

In sports wagering markets the patterns of team streak performance that bettors respond to cannot be assumed away as in the casino and lottery context. Because hot hand beliefs may be justifiable in these settings, any analysis must investigate patterns in market betting behavior relative to patterns in actual team performance. Camerer (1989) showed that for basketball teams currently on a winning streak, the point spread in the actual game (the difference in points scored between two teams) exceeded the final market point spread (the "beliefs" of the betting market about the point spread) less than half the time, demonstrating that the market is biased and some bettors believe that teams who have recently won will exceed the spread more than they actually do. Nevertheless, Camerer found this bias to be too small to be profitably exploited. Brown and Sauer (1993) conducted a study confirming the analysis of Camerer, but finding that the data do not provide evidence that believing in a hot hand is a fallacy in this setting. Brown and Sauer analyze basketball betting market data in a way that jointly estimates hot hand beliefs and the hot hand effect. They find evidence of modest hot hand beliefs in market aggregates, but they cannot rule out the possibility that these beliefs are based on a real hot hand effect in team performance.[83] In a data set consisting of seven years of basketball point spreads, Paul and Weinbach (2005) find that for certain minimum win streak lengths, a betting strategy, which bets against the teams on a streak, will exceed chance levels and be borderline profitable, over the course of seven years. More recently, Arkes (2011) has found some suggestive evidence of a hot hand effect in basketball team performance, in that realized point spreads appear to be sensitive to variability in recent success, over and above many alternative measures of the strength of the two teams playing.[84] It appears that sports bettors are willing to accept odds that are less than fair for bets on teams that are

[82] See Yuan et al. (2014) for more recent and clear evidence that, in order to play the numbers that "lucky" players have selected, people are willing to accept a significantly lower payout in the (unlikely) state of the world in which they win the lottery.

[83] Using a different approach and a different data set (football betting markets), Avery and Chevalier (1999) find that hot hand beliefs influence movements in the market point spread from the initial spread set by odds-makers to the final spread set after their interaction with the bettors. The study was intended to identify bettor sentiment and cannot determine if the hot hand belief has an excessive influence on betting markets, or if odds makers do not believe enough in the hot hand effect when setting the opening betting line. Durham et al. (2005) also analyze movements in point spreads in football betting markets, but find evidence consistent with the gambler's fallacy rather than the hot hand fallacy.

[84] Using a seemingly unrelated regression model with the regression equation specifications assumption that the factors determining the market point spread and the realized point spread were identical, as in Brown and Sauer (1993), Arkes (2011) found the market point spread to be more sensitive to a team's hot streak than the realized point spread.

currently on a winning streak, but bettors are correct that teams on a streak are more likely to win, and the biases in the odds are too small to be profitably exploited.[85] Further, as the beliefs of bettors are not measured directly, the mechanism that produces this phenomenon may not be a hot hand belief, but instead may be the salience of the usual media coverage and sports commentator opinion that accompanies a team that is performing well.

Finally, in contrast to the test of hot hand fallacy in the basketball domain, these field studies by and large all study the behavior of amateurs rather than experts, with less opportunity to learn from expert feedback, training, and advice. For example, amateur gamblers, bettors, and lottery players are generally less exposed to the type of advice, direction, and feedback to which an expert basketball player is exposed; therefore, even if a persistent bias were observed among them (and were costly), it could not be considered as indicative of a deeply rooted cognitive illusion. Further, evidence in these domains can hardly be said to be representative of decision making more generally, because many of the participants likely view their behavior as recreational, and for those who do not, they are *gamblers*, and not necessarily representative of the population at large.

Lab evidence of the hot hand fallacy

A large body of early lab experiments have documented the systematic biases present in the human perception of randomly generated sequences, including behavior consistent with the hot hand fallacy.[86] These experiments, however, are subject to several important limitations related to ex-

[85] In American college football wagering markets Sinkey and Logan (2013) find that betting houses correctly set point spreads to account for the a hot hand effect in team performance, i.e. point spreads are fair and there is no evidence of a hot hand fallacy in college football betting markets. In American professional football wagering markets, Lee and Smith (2002) document a related "regression fallacy", and find that participants over-weight past performance against the spread. A betting strategy that bets on each game and places money on the team that in net has performed worse relative to the spread in previous games (and bets in proportion to how much worse this team has performed), will win at greater than chance levels in 5 out of 8 seasons (and overall). Moreover, this strategy is profitable in 4 out of 8 individual seasons. This evidence indicates that either odds-makers overweight observed performance against the spread when setting the line, leaving themselves moderately exposed but not exploited over the 8 seasons, or that the line was set to clear markets, and that bettors overweight this information. Because performance against the spread is a salient and relevant piece of information, and bettors have structural uncertainty, it is difficult to know what its information content is and whether bettors are behaving irrationally.

[86] For an early review see Wagenaar (1972). For more recent reviews see Bar-Hillel and Wagenaar (1991); Nickerson (2002); Oskarsson, Boven, McClelland, and Hastie (2009); Rabin (2002). There is also recent work providing evidence that humans (and other primates) perform better in a prediction task involving a sequence with an unknown data generating process if the sequence they face is positively autocorrelated rather than negatively autocorrelated; further, in these studies people prefer to predict from a positively autocorrelated sequence rather than from an uncorrelated sequence, yet they prefer to predict from an uncorrelated sequence rather than from a negatively autocorrelated sequence (Blanchard, Wilke, and Hayden 2014; Scheibehenne, Wilke, and Todd 2011). For recent attempts to explain (and rationalize) these biases see Caruso, Waytz, and Epley (2010); Hahn and Warren (2009); Sun and Wang (2010); Whitson and Galinsky (2008).

ternal validity, such as the role of amateur participants, experimenter demand effects, and lack of incentives (Einhorn and Hogarth 1981; Lopes 1982; Morrison and Ordeshook 1975). While GVT's basketball study was intended to address many of the limitations present in earlier studies, there is also a body of more recent laboratory studies, which focus on financial applications and theory testing, conduct incentivized experiments, and which can, in principle, provide clearer evidence of the hot hand fallacy than is possible with field data. The designs are varied, and involve subjects determining the probability that one data generating process, rather than another, produced an observed outcome (i.e. signal detection) (Massey and Wu 2005; Offerman and Sonnemans 2004), first observing several periods of a random walk process, and then determining the probability that the next (binary) outcome is a success (Asparouhova, Hertzel, and Lemmon 2009; Bloomfield and Hales 2002), or, similarly, predicting whether the next outcome will be a success, but only after first being given the option to pay for possibly useless prediction advice (Huber, Kirchler, and Stöckl 2010; Powdthavee and Riyanto 2012).[87] In general, aside from providing proper incentives, the same limitations noted with respect to early laboratory studies also arise in these newer studies.[88] Results are sometimes consistent with the hot hand fallacy and other times not: Massey and Wu (2005) find that subjects generally over-react to signals of possible regime shifts when predicting the next outcome, but under-react to the same signals when giving a probability assessment (p) that a regime switch has already occurred. While Asparouhova et al. (2009) observe a pattern of behavior that is generally consistent with the predictions of a model by Rabin and Vayanos (2010), subjects do not know what the data generating process of observed outcome sequences is (are), so any beliefs are rationalizable (thus none can be said to be fallacious).[89] Huber et al. (2010) find that subjects are willing to pay for advice from an "expert," regarding what the next outcome of a computerized coin flip will be, particularly after the expert's predictions have been correct for many periods in a row. Because the mechanism behind the computerized coin toss is opaque, and subjects are not informed as to how the expert makes predictions, their behavior can easily be

[87] See Dohmen, Falk, Huffman, Marklein, and Sunde (2009) for a survey linking the tendency to believe in the hot hand with long-term unemployment.

[88] However, quadratic scoring (and other) payment mechanisms employed often suffer from a "flat-max" problem (Keren 1991).

[89] In a second experiment the authors inform the subjects that the data generating process is a sequence of flips from a fair coin, and in this case subjects' beliefs become more gambler's fallacy intensive (consistent with Rabin and Vayanos (2010)) but also anticipate positive recency for *shorter* streak lengths, which is not consistent with the model's predictions. A field study of college football betting markets Durham et al. (2005) find almost an opposite pattern of results.

viewed as a rational response to the presence of structural uncertainty, and therefore it is difficult to judge their decisions as fallacious. Further, as subjects observe more predictions and outcomes over time they purchase predictions considerably less often, which suggests that they rationally adapt as they acquire more information about the environment. Powdthavee and Riyanto (2012) run a similar experiment, but with a design that makes it transparent to subjects that experts' advice is of no value, and they observe the same general pattern of behavior as in Huber et al. (2010). Nevertheless, as is common in earlier studies, there remains a distinct possibility that observed behavior is caused by an experimenter demand effect: the task involves subjects making decisions in a spare and repetitive environment, while being provided information that is highly salient, varies, and yet is irrelevant.[90] In fact, the results are shown to not be robust to small changes in the quantity of available advice: in a follow-up study, Foster, Frijters, Powdthavee, and Riyanto (2014) find that by merely providing additional useless "expert" predictions the effect is eliminated (albeit in a different subject pool).

Overall, we asses laboratory evidence to be rich and informative in its ability to reveal basic behavioral tendencies, but not able to demonstrate that the hot hand fallacy is either a massive or widespread cognitive illusion—both of which were seemingly demonstrated by evidence from the basketball domain. In particular, in the laboratory the observed effects have been mixed, sensitive to minor changes in context, and have not been demonstrated to be robust to learning, feedback, or decision making by non-amateurs.

5.3 Hot hand effect more generally

Part of the fame of the hot hand fallacy is due to how counter-intuitive the result is; consistent with the beliefs of expert players and coaches, it is natural to expect professionals to sometimes enter into relatively superior performance states. In fact, not only does our finding of a hot hand effect in basketball shooting accord with intuition, but it also agrees with existing literature on human performance more generally. In particular, when considering individual performance, whether in a motor task, or an intellectual task, it seems unsurprising that changes in states, such as expectancy and confidence (Bandura 1982; Damisch, Stoberock, and Mussweiler 2010; Nelson and Furst 1972),

[90] Powdthavee and Riyanto (2012) consider this possibility of demand effects, positing that subjects may think: "I know that predictions contained within these envelopes are useless. But if they are really useless, then why would they be here in the first place?"

attention and flow (Csikszentmihalyi 1988; Kahneman 1973), or neural noise (Churchland et al. 2006) can lead to clusters of altered performance. With the detection of the hot hand effect in our study, and in every extant controlled shooting study, it becomes natural to consider the degree to which the hot hand effect is also a general phenomenon in a broader class of performance environments.

In their seminal study, GVT were careful to restrict their conclusion that there is no hot hand effect to the domain of basketball shooting, in which belief in the effect was already known to be near-unanimous. The pairing of these facts exhibited the strength of the hot hand fallacy, and suggested its relevance to decision making more generally. (Alter and Oppenheimer 2006; Gilovich et al. 1985; Tversky and Gilovich 1989a,b). Nevertheless, because the possibility that highly motivated professionals have a tendency to exhibit performance clustering is also of general interest, a large and separate related literature has emerged from the original GVT study focusing on the existence hot hand performance, independent of the beliefs of decision makers. While most of these studies pertain to sport (Bar-Eli et al. 2006), other domains are also studied, including financial management (Hendricks et al. 1993). The overall evidence for a hot hand effect has been mixed (Avugos et al. 2013b; Carhart 1997), but several studies have found evidence of possible hot hand performance among hedge fund managers (Jagannathan et al. 2010) and in the sport domains of professional tennis (Klaassen and Magnus 2001), bowling (Dorsey-Palmateer and Smith 2004; Yaari and David 2012), horseshoes (Smith 2003), golf (Connolly and Rendleman 2008), darts and golf-putting (Gilden and Wilson 1995), and more recently, and prominently, in baseball (Green and Zwiebel 2013).[91,92] While these studies do share some of the limitations discussed in Section 2.1—e.g. the pooling of data over extended time periods, and the inability to separate a hot hand from a cold hand—they are nevertheless instructive, as they find evidence in the direction consistent with the hot hand effect.[93,94]

That our discovery of the hot hand effect in its canonical domain accords with both intuition,

[91] The existence of streaks in professional baseball has been a topic of contention in the statistics literature; see Albert (1993, 2008, 2013); Albert and Williamson (2001); Albright (1993a,b); McCotter (2010); Stern and Morris (1993). Also see Bock, Maewal, and Gough (2012); Gray and Allsop (2013).

[92] Raab, Gula, and Gigerenzer (2011) find evidence of a hot hand in volleyball, though the validity issues discussed in Section 2.1 apply, as volleyball is a team sport.

[93] Green and Zwiebel (2013) are an exception in that they separately identify hot hand and cold hand effects.

[94] There is also some concern that variable incentives may drive some of these effects, especially in tournaments—contest incentives can induce performance clustering (McFall, Knoeber, and Thurman 2009; Szymanski 2003).

and existing literature from other domains, further dismisses the notion of hot hand belief as necessarily a fallacy. Indeed, coaches of sports teams who must allocate the ball, managers of companies who must allocate resources, and researchers choosing how to allocate their effort across time can now feel more confident that, on occasion, they may be in the midst of, or observing, a period of unusually superior performance.[95]

6 Conclusion

We test for the hot hand fallacy in the canonical domain of basketball, where the supporting evidence has long been considered the strongest, and most readily generalizable to other domains. Using a novel empirical strategy to study the best available basketball shooting data—including that from our own field experiment and the original study of Gilovich, Vallone, and Tversky (1985)—we provide convincing evidence that contrary to nearly 30 years of research, the belief in the hot hand is not a fallacy, and that, surprisingly, a re-analysis of the original evidence supports this conclusion. Further, our review of the hot hand fallacy literature—across a variety of economic decision making environments—provides no indication that the fallacy exists in domains in which there is a level of expertise, incentives, or opportunities for learning comparable to what is present in the canonical basketball domain. This conclusion places a sharp bound on the relevance of the hot hand fallacy for economic decision making more generally.

Our results establish that expert players and coaches are justified in believing in the hot hand, and the elicited beliefs of our expert players suggest that they are able to identify the shooters who have a greater tendency to become hot. Nevertheless, a formal test of the strength of beliefs in amateurs, in addition to experts, is desirable, as it could provide a clearer answer to the question of whether people at times perceive a hot hand when it is not present, or perceive the effect to be excessively strong when it is present, and whether non-experts have a greater tendency to make these types of mistakes. While it is reasonable to suspect that some people will over-infer based on the limited information that they receive (Barberis et al. 1998; Burns 2004; Mullainathan 2002; Rabin 2002; Rabin and Vayanos 2010), a substantial level of over-inference would be surprising among highly incentivized experts, given that it would represent a stark deviation from the roughly

[95] There is some recent evidence that momentum in performance influences organizational risk-taking (Lehman and Hahn 2013).

optimal shooting decisions that have been observed in most situations (Goldman and Rao 2014).[96] We leave for future work the important task of further characterizing in which settings, and by which individuals (expert or amateur), the strength of hot hand beliefs are well-calibrated, and in which settings they are not.

[96]We suggest three reasons why it may be reasonable to expect experts to over-weight recent success beyond what can be justified by its ability to predict subsequent performance: (1) in a world in which people are unsure if there has been a change in a player's probability of success (regime shift), the gambler's fallacy—a mistake for which the evidence is considerably stronger—implies that people will over-infer from recent success (Rabin 2002; Rabin and Vayanos 2010), (2) in a world in which coaches and players are presented with more information relating to a player's physical and mental state than can be attended to, yet must nevertheless assess the player's current success probability without knowing which information is maximally diagnostic, bounded rationality may make it adaptive to pay special attention to salient streaks and recent shooting performance (see Burns (2004) for a discussion of why this behavior can be adaptive, even in the absence of hot hand shooting), (3) humans (and other primates) may have evolved a tendency to expect positive autocorrelation from sequential data because it is adaptive in a world in which resources have a tendency to cluster together (Blanchard et al. 2014; Fawcett, Fallenstein, Higginson, Houston, Mallpress, Trimmer, and McNamara 2014; Scheibehenne et al. 2011; Wilke and Barrett 2009).

References

Aharoni, G. and O. H. Sarig (2011): "Hot hands and equilibrium," *Applied Economics*, 44, 2309–2320.

Albert, J. (1993): "Comment on "A Statistical Analysis of Hitting Streaks in Baseball" by S. C. Albright," *Journal of the American Statistical Association*, 88, 1184–1188.

——— (2008): "Streaky Hitting in Baseball," *Journal of Quantitative Analysis in Sports*, 4, 1–32.

——— (2013): "Looking at spacings to assess streakiness," *Journal of Quantitative Analysis in Sports*.

Albert, J. and P. Williamson (2001): "Using Model/Data Simulations to Detect Streakiness," *The American Statistician*, 55, 41–50.

Albright, S. C. (1993a): "Rejoinder," *Journal of the American Statistical Association*, 88, 1194–1196.

——— (1993b): "A Statistical Analysis of Hitting Streaks in Baseball," *Journal of the American Statistical Association*, 88, 1175–1183.

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

Angrist, J. D. and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion.*

Arkes, J. (2010): "Revisiting the Hot Hand Theory with Free Throw Data in a Multivariate Framework," *Journal of Quantitative Analysis in Sports*, 6.

——— (2011): "Do Gamblers Correctly Price Momentum in NBA Betting Markets?" *Journal of Prediction Markets*, 5, 31–50.

——— (2013): "Misses in 'Hot Hand' Research," *Journal of Sports Economics*, 14, 401–410.

Asparouhova, E., M. Hertzel, and M. Lemmon (2009): "Inference from Streaks in Random Outcomes: Experimental Evidence on Beliefs in Regime Shifting and the Law of Small Numbers," *Management Science*, 55, 1766–1782.

Attali, Y. (2013): "Perceived Hotness Affects Behavior of Basketball Players and Coaches," *Psychological Science*, forthcoming.

Avery, C. and J. Chevalier (1999): "Identifying Investor Sentiment from Price Paths: The Case of Football Betting," *Journal of Business*, 72, 493–521.

Avugos, S., M. Bar-Eli, I. Ritov, and E. Sher (2013a): "The elusive reality of efficacy performance cycles in basketball shooting: analysis of players performance under invariant conditions," *International Journal of Sport and Exercise Psychology*, 11, 184–202.

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

Ayton, P. and I. Fischer (2004): "The hot hand fallacy and the gamblers fallacy: Two faces of subjective randomness?" *Memory & Cognition*, 21, 1369–1378.

Ball, C. T. (2012): "Not all streaks are the same: Individual differences in risk preferences during runs of gains and losses." *Judgement and Decision Making*, 7, 452–461.

Bandura, A. (1982): "Self-Efficacy Mechanism in Human Agency," *American Psychologist*, 37, 122–147.

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

Bar-Hillel, M. and W. A. Wagenaar (1991): "The perception of randomness," *Advances in Applied Mathematics*, 12, 428–454.

Barber, B. M. and T. Odean (2008): "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *The Review of Financial Studies*, 21, 785–818.

Barber, B. M., T. Odean, and N. Zhu (2009a): "Do Retail Trades Move Markets?" *The Review of Financial Studies*, 22, 151–186.

——— (2009b): "Systematic noise," *Journal of Financial Markets*, 12, 547–569.

Barberis, N., A. Shleifer, and R. Vishny (1998): "A Model of Investor Sentiment," *Journal of Financial Economics*, 49, 307–343.

Barberis, N. and R. Thaler (2003): "A survey of behavioral finance," *Handbook of the Economics of Finance*, 1, 1053–1128.

Benartzi, S. (2001): "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," *Journal of Finance*, 56, 1747–1764.

Benartzi, S. and R. H. Thaler (2007): "Heuristics and Biases in Retirement Savings Behavior," *Journal of Economic Perspectives*, 21, 81–104.

Berk, J. B. and R. Green (2004): "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112, 1269–1295.

Blanchard, T. C., A. Wilke, and B. Y. Hayden (2014): "Hot-hand bias in rhesus monkeys." *Journal of Experimental Psychology: Animal Learning and Cognition*, 40, 280–.

Bloomfield, R. and J. Hales (2002): "Predicting the next step of a random walk: experimental evidence of regime-shifting beliefs," *Journal of Financial Economics*, 65, 397–414.

Bock, J. R., A. Maewal, and D. A. Gough (2012): "Hitting Is Contagious in Baseball: Evidence from Long Hitting Streaks," *PLoS One*, 7, 12.

Bocskocsky, A., J. Ezekowitz, and C. Stein (2014): "The Hot Hand: A New Approach to an Old 'Fallacy'," 8th Annual Mit Sloan Sports Analytics Conference.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012): "Salience Theory of Choice Under Risk," *Quarterly Journal of Economics*, 127, 1243–1285.

Brav, A., J. Heaton, and A. Rosenberg (2004): "The rational-behavioral debate in financial economics," *Journal of Economic Methodology*, 11, 393–409.

Brav, A. and J. B. Heaton (2002): "Competing Theories of Financial Anomalies," *The Review of Financial Studies*, 15, 575–606.

Brown, W. A. and R. D. Sauer (1993): "Does the Basketball Market Believe in the Hot Hand? Comment," *American Economic Review*, 83, 1377–1386.

Burns, B. D. (2004): "Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand"," *Cognitive Psychology*, 48, 295–331.

Camerer, C. F. (1989): "Does the Basketball Market Believe in the 'Hot Hand,'?" *American Economic Review*, 79, 1257–1261.

Cao, Z. (2011): "Essays on Behavioral Economics," Ph.D. thesis, Oregon State University.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Carlson, K. A. and S. B. Shu (2007): "The rule of three: how the third event signals the emergence of a streak," *Organizational Behavior and Human Decision Processes*, 104, 113–121.

Caruso, E. M., A. Waytz, and N. Epley (2010): "The intentional mind and the hot hand: Perceiving intentions makes streaks seem likely to continue," *Cognition*, 116, 149–153.

Case, K. E. and R. J. Shiller (1988): "The behavior of home buyers in boom and post-boom markets," *New England Economic Review*, 29–46.

Case, K. E., R. J. Shiller, and A. Thompson (2012): "What have they been thinking? Home buyer behavior in hot and cold markets," Tech. rep., National Bureau of Economic Research.

Chevalier, J. and G. Ellison (1997): "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 105, 1167–1200.

——— (1999): "Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance," *Journal of Finance*, 54, 875–899.

Churchland, M. M., A. Afshar, and K. V. Shenoy (2006): "A Central Source of Movement Variability," *Neuron*, 52, 1085–1096.

Clotfelter, C. and P. Cook (1991): "Lotteries in the real world," *Journal of Risk and Uncertainty*, 4, 227–232.

Clotfelter, C. T. and P. J. Cook (1993): "The 'Gambler's Fallacy' in Lottery Play," *Management Science*, 39, 1521–1525.

CONNOLLY, R. A. AND R. J. RENDLEMAN, JR. (2008): "Skill, Luck, and Streaky Play on the PGA Tour," *Journal of the American Statistical Association*, 103, 74–88.

CROSON, R. AND J. SUNDALI (2005): "The Gamblers Fallacy and the Hot Hand: Empirical Data from Casinos," *Journal of Risk and Uncertainty*, 30, 195–209.

CSIKSZENTMIHALYI, M. (1988): "The flow experience and its significance for human psychology," in *Optimal experience: Psychological studies of flow in consciousness*, ed. by M. Csikszentmihalyi and I. S. Csikszentmihalyi, New York, NY, US: Cambridge University, chap. 2.

DAMISCH, L., B. STOBEROCK, AND T. MUSSWEILER (2010): "Keep Your Fingers Crossed!: How Superstition Improves Performance," *Psychological Science*, 21, 1014–1020.

DE BONDT, W. F. (1998): "A portrait of the individual investor," *European Economic Review*, 42, 831 – 844.

DE BONDT, W. P. (1993): "Betting on trends: Intuitive forecasts of financial risk and return," *International Journal of Forecasting*, 9, 355–371.

DE LONG, J. B., A. SHLEIFER, L. H. SUMMERS, AND R. J. WALDMANN (1990): "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance*, 45, 379–395.

——— (1991): "The Survival of Noise Traders In Financial-markets," *Journal of Business*, 64, 1–19.

DELLAVIGNA, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

DIXIT, A. K. AND B. J. NALEBUFF (1991): *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*, W.W. Norton & Company.

DOHMEN, T., A. FALK, D. HUFFMAN, F. MARKLEIN, AND U. SUNDE (2009): "Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample," *Journal of Economic Behavior & Organization*, 72, 903–915.

DORSEY-PALMATEER, R. AND G. SMITH (2004): "Bowlers' Hot Hands," *The American Statistician*, 58, 38–45.

DURHAM, G. R., M. G. HERTZEL, AND J. S. MARTIN (2005): "The Market Impact of Trends and Sequences in Performance: New Evidence," *Journal of Finance*, 60, 2551–2569.

EINHORN, H. J. AND R. M. HOGARTH (1981): "Behavioral decision theory: Processes of judgment and choice," 19, 1–31.

ERNST, M. D. (2004): "Permutation Methods: A Basis for Exact Inference," *Statistical Science*, 19, 676–685.

FALK, R. AND C. KONOLD (1997): "Making sense of randomness implicit encoding as a basis for judgment," *Psychological Review*, 104, 301–318.

Fama, E. F. (1998): "Market efficiency, long-term returns, and behavioral finance," *Journal of Financial Economics*, 49, 283 – 306.

Fama, E. F. and K. R. French (2010): "Luck versus Skill in the Cross-Section of Mutual Fund Returns," *Journal of Finance*, 65, 1915–1947.

——— (2012): "Size, value,and momentum in international stock returns," *Journal of Financial Economics*, 105, 457–472.

Fawcett, T. W., B. Fallenstein, A. D. Higginson, A. I. Houston, D. E. W. Mallpress, P. C. Trimmer, and J. M. McNamara (2014): "The evolution of decision rules in complex environments," *Trends in Cognitive Sciences*, 18, 153–161.

Foster, G., P. Frijters, N. Powdthavee, and Y. E. Riyanto (2014): "Is Good Fortune a Tradable Asset? Further Experimental Evidence on the Economic Value of Luck," Conference Presentation: SABE, Lake Tahoe, 2014.

French, K. (2008): "The Cost of Active Investing," *Journal of Finance*, 63, 1537–1573.

Galbo-Jørgensen, C. B., S. Suetens, and J.-R. Tyran (2013): "Predicting Lotto Numbers A natural experiment on the gamblers fallacy and the hot hand fallacy," Working Paper.

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

Gilovich, T., R. Vallone, and A. Tversky (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 17, 295–314.

Goldman, M. and J. M. Rao (2014): "Loss Aversion and Pressure in Highly Experienced Agents," Working Paper.

Good, P. (2005): *Permutation, Parametric, and Bootstrap Tests of Hypotheses.*, Springer.

Gould, S. J. (1989): "The streak of streaks," *Chance*, 2, 10–16.

Gray, R. and J. Allsop (2013): "Interactions between performance pressure, performance streaks, and attentional focus." *Journal of Sport and Exercise Psychology*, 35, 368–386.

Green, B. S. and J. Zwiebel (2013): "The Hot Hand Fallacy: Cognitive Mistakes or Equilibrium Adjustments?" Working Paper.

Guryan, J. and M. S. Kearney (2008): "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," *American Economic Review*, 98, 458–473.

Hahn, U. and P. A. Warren (2009): "Perceptions of randomness: why three heads are better than four," *Psychological Review*, 116, 454–461.

Hakes, J. K. and R. D. Sauer (2006): "An Economic Evaluation of the Moneyball Hypothesis," *Journal of Economic Perspectives*, 20, 173–186.

Harrison, G. W. and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42, 1009–1055.

HENDRICKS, D., J. PATEL, AND R. ZECKHAUSER (1993): "Hot hands in mutual funds: Short-run persistence of relative performance," *Journal of Finance*, 48, 93–130.

HOOKE, R. (1989): "Basketball, baseball, and the null hypothesis," *Chance*, 2, 35–37.

HOUTHAKKER, H. S. (1961): "Systematic and Random Elements in Short-Term Price Movements," *American Economic Review*, 51, 164–172.

HUBER, J., M. KIRCHLER, AND T. STÖCKL (2010): "The hot hand belief and the gambler's fallacy in investment decisions under risk," *Theory and Decision*, 68, 445–462.

JAGACINSKI, R. J., K. M. NEWELL, AND P. D. ISAAC (1979): "Predicting the Success of a Basketball Shot at Various Stages of Execution," *Journal of Sport Psychology*, 1, 301–310.

JAGANNATHAN, R., A. MALAKHOV, AND D. NOVIKOV (2010): "Do Hot Hands Exist among Hedge Fund Managers? An Empirical Evaluation," *Journal of Finance*, 65, 217–255.

JEGADEESH, N. AND S. TITMAN (1993): "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48, 65–91.

KAHNEMAN, D. (1973): *Attention and Effort.*, Prentice Hall.

——— (2011): *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

KAHNEMAN, D. AND M. W. RIEPE (1998): "Aspects of Investor Psychology: Beliefs, preferences, and biases investment advisors should know about," *Journal of Portfolio Management*, 24, 1–21.

KEARNEY, M. S. (2005): "State lotteries and consumer behavior," *Journal of Public Economics*, 89, 2269–2299.

KEREN, G. (1991): "Calibration and probability judgements: Conceptual and methodological issues," *Acta Psychologica*, 77, 217–273.

KEREN, G. B. AND W. A. WAGENAAR (1985): "On the psychology of playing blackjack: Normative and descriptive considerations with implications for decision theory." *Journal of Experimental Psychology: General*, 114, 133–158.

KLAASSEN, F. J. G. M. AND J. R. MAGNUS (2001): "Are Points in Tennis Independent and Identically Distributed? Evidence From a Dynamic Binary Panel Data Model," *Journal of the American Statistical Association*, 96, 500–509.

KOEHLER, J. J. AND C. A. CONLEY (2003): "The "hot hand" myth in professional basketball," *Journal of Sport and Exercise Psychology*, 25, 253–259.

KORB, K. B. AND M. STILLWELL (2003): "The Story of The Hot Hand: Powerful Myth or Powerless Critique?" Working Paper.

KUHN, P., P. KOOREMAN, A. SOETEVENT, AND A. KAPTEYN (2011): "The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery," *American Economic Review*, 101, 2226–2247.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994): "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance*, 49, 1541–1578.

Larkey, P. D., R. A. Smith, and J. B. Kadane (1989): "Its Okay to Believe in the Hot Hand," *Chance*, 2, 22–30.

Lee, M. and G. Smith (2002): "Regression to the mean and football wagers," *Journal of Behavioral Decision Making*, 15, 329–342.

Lehman, D. W. and J. Hahn (2013): "Momentum and organizational risk taking: Evidence from the National Football League." *Management Science*, 59, 852–868.

LeRoy, S. F. (2004): "Rational Exuberance," *Journal of Economic Literature*, 42, 783–804.

Lewellen, J. (2002): "Momentum and Autocorrelation in Stock Returns," *The Review of Financial Studies*, 15, 533–564.

Lewis, M. (2004): *Moneyball: The Art of Winning an Unfair Game*, W. W. Norton & Company.

List, J. A. (2003): "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118, 41–71.

Loh, R. K. and M. Warachka (2012): "Streaks in Earnings Surprises and the Cross-Section of Stock Returns," *Management Science*, 58, 1305–1321.

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

Malkiel, B. G. (2011): *A random walk down Wall Street: the time-tested strategy for sucessful investing*, New York: W. W. Norton & Company.

Massey, C. and G. Wu (2005): "Detecting Regime Shifts: The Causes of under- and Overreaction," *Management Science*, 51, 932–947.

McCotter, T. (2010): "Hitting Streaks Dont Obey Your Rules: Evidence that hitting streaks are not just byproducts of random variation," *Chance*, 23, 52–57.

McFall, T. A., C. R. Knoeber, and W. N. Thurman (2009): "Contests, Grand Prizes, and the Hot Hand," *Journal of Sports Economics*, 10, 236–255.

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

Mood, A. M. (1940): "The Distribution Theory of Runs," *The Annals of Mathematical Statistics*, 11, 367–392.

Morrison, R. S. and P. C. Ordeshook (1975): "Rational choice, light guessing and the gambler's fallacy," *Public Choice*, 22, 79–89–.

Mullainathan, S. (2002): "Thinking Through Categories," Working Paper.

Narayanan, S. and P. Manchanda (2012): "An empirical analysis of individual level casino gambling behavior," 10, 27–62.

Neiman, T. and Y. Loewenstein (2011): "Reinforcement learning in professional basketball players," *Nature Communications*, 2:569.

Nelson, L. R. and M. L. Furst (1972): "An Objective Study of the Effects of Expectation on Competitive Performance," *Journal of Psychology*, 81, 69–72.

Nickerson, R. S. (2002): "The production and perception of randomness," *Psychological Review*, 109, 350–357.

Odean, T. (1999): "Do Investors Trade Too Much?" *American Economic Review*, 89, 1279–1298.

Offerman, T. and J. Sonnemans (2004): "What's Causing Overreaction? An Experimental Investigation of Recency and the Hot-hand Effect," *The Scandinavian Journal of Economics*, 106, 533–554.

Oskarsson, A. T., L. V. Boven, G. H. McClelland, and R. Hastie (2009): "Whats next? Judging sequences of binary events," *Psychological Bulletin*, 135, 262–385.

Padoa-Schioppa, C. and A. Rustichini (2014): "Rational Attention and Adaptive Coding: A Puzzle and a Solution," *American Economic Review*, 104, 507–13.

Pástor, L. and P. Veronesi (2009): "Learning in financial markets," Working Paper.

Paul, R. J. and A. P. Weinbach (2005): "Bettor Misperceptions in the NBA: The Overbetting of Large Favorites and the 'Hot Hand'," *Journal of Sports Economics*, 6, 390–400.

Powdthavee, N. and Y. E. Riyanto (2012): "Would you pay for transparently useless advice? A test of boundaries of beliefs in the folly of predictions," *The Review of Economics and Statistics*, forthcoming, working paper.

Raab, M., B. Gula, and G. Gigerenzer (2011): "The Hot Hand Exists in Volleyball and Is Used for Allocation Decisions," *Journal of Experimental Psychology: Applied*, 18, 81–94.

Rabin, M. (2002): "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 117, 775–816.

Rabin, M. and D. Vayanos (2010): "The Gamblers and Hot-Hand Fallacies: Theory and Applications," *Review of Economic Studies*, 77, 730–778.

Rao, J. M. (2009a): "Experts' Perceptions of Autocorrelation: The Hot Hand Fallacy Among Professional Basketball Players," Working Paper.

——— (2009b): "When the Gambler's Fallacy becomes the Hot Hand Fallacy: An Experiment with Experts and Novices," Working Paper.

Reifman, A. (2012): *Hot Hand: The Statistics Behind Sports' Greatest Streaks*, Potomac Books.

Russell, W. F. (1969): "I'm Not Involved Anymore," *Sports Illustrated*, August 04.

Sauer, R. D. (1998): “The Economics of Wagering Markets,” *Journal of Economic Literature*, 36, 2021–2064.

Scheibehenne, B., A. Wilke, and P. M. Todd (2011): “Expectations of clumpy resources influence predictions of sequential events,” *Evolution and Human Behavior*, 32, 326 – 333.

Sinkey, M. and T. Logan (2013): “Does the Hot Hand Drive the Market?” *Eastern Economic Journal*, Advance online publication, doi:10.1057/eej.2013.33.

Sirri, E. R. and P. Tufano (1998): “Costly Search and Mutual Fund Flows,” *Journal of Finance*, 53, 1589–1622.

Smith, G. (2003): “Horseshoe pitchers hot hands,” *Psychonomic Bulletin and Review*, 10, 753–758.

Smith, G., M. Levere, and R. Kurtzman (2009): “Poker Player Behavior After Big Wins and Big Losses,” *Management Science*, 55, 1547–1555.

Stern, H. S. and C. N. Morris (1993): “Comment on “A Statistical Analysis of Hitting Streaks in Baseball” by S. C. Albright,” *Journal of the American Statistical Association*, 88, 1189–1194.

Stewart, N., S. Reimers, and A. J. L. Harris (2014): “On the Origin of Utility, Weighting, and Discounting Functions: How They Get Their Shapes and How to Change Their Shapes,” *Management Science*, forthcoming.

Stone, D. F. (2012): “Measurement error and the hot hand,” *The American Statistician*, 66, 61–66, working paper.

Sun, Y. and H. Wang (2010): “Gamblers fallacy, hot hand belief, and the time of patterns,” *Judgement and Decision Making*, 5, 124–132.

Sundali, J. and R. Croson (2006): “Biases in casino betting: The hot and the gamblers fallacy,” *Judgement and Decision Making*, 1, 1–12.

Swartz, T. (1990): “Letter to the editor: More on the “hot hand”,” *Chance*, 3, 6–7.

Szymanski, S. (2003): “The economic design of sporting contests,” *Journal of Economic Literature*, 41, 1137–1187.

Terrell, D. (1994): “A test of the gambler’s fallacy: Evidence from pari-mutuel games,” *Journal of Risk and Uncertainty*, 8, 309–317.

Thaler, R. H. and E. J. Johnson (1990): “Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice,” *Management Science*, 36, 643–660.

Thaler, R. H. and C. R. Sunstein (2008): *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press.

Thaler, R. H. and W. T. Ziemba (1988): “Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries,” *Journal of Economic Perspectives*, 2, 161–174.

Tversky, A. and T. Gilovich (1989a): “The cold facts about the “hot hand” in basketball,” *Chance*, 2, 16–21.

——— (1989b): "The"Hot Hand": Statistical Reality or Cognitive Illusion?" *Chance*, 2, 31–34.

Tversky, A. and D. Kahneman (1974): "Judgment under Uncertainty: Heuristics and Biases," *Science*, 185, 1124–1131.

Wagenaar, W. A. (1972): "Generation of Random Sequences by Human Individuals: A Critical Survey of the Literature," *Psychological Bulletin*, 77, 65–72.

Wardrop, R. L. (1999): "Statistical Tests for the Hot-Hand in Basketball in a Controlled Setting," Working paper, University of Wisconsin - Madison.

Whitson, J. A. and A. D. Galinsky (2008): "Lacking Control Increases Illusory Pattern Perception," *Science*, 322, 115–117.

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

Xu, J. and N. Harvey (2014): "Carry on winning: The gambler's fallacy creates hot hand effects in online gambling," *Cognition*, 131, 173 – 180.

Yaari, G. and G. David (2012): ""Hot Hand" on Strike: Bowling Data Indicates Correlation to Recent Past Results, Not Causality," *PLoS One*, 7, 1–9.

Yaari, G. and S. Eisenmann (2011): "The Hot (Invisible?) Hand: Can Time Sequence Patterns of Success/Failure in Sports Be Modeled as Repeated Random Independent Trials?" *PLoS One*, 6, 1–10.

Yuan, J., G.-Z. Sun, and R. Siu (2014): "The Lure of Illusory Luck: How Much Are People Willing to Pay for Random Shocks," *Journal of Economic Behavior & Organization*, forthcoming.

A Appendix: Experimental Procedures & Instructions

A.1 Instructions: Shooting Experiment

Shooting

INSTRUCTIONS: SHOOTERS

Your Task:

You will shoot 300 times with your toes just behind the line marked with tape point to line. There is no need to move from the line because the rebounder will rebound each of your shots and pass the ball back to you once I signal him to do so via a clear audible signal. Once you have received the ball you can shoot when you like.

As you shoot I will be sitting on the side of the court recording whether you make or miss each shot. The video cameras you see will be filming your shots.

Payoffs:

Of your 300 shot opportunities 10 of them have been selected at random as shots for which you will be paid. For each of these 10 selected shots that you make you will be paid 6,00 Euros. For each of these 10 selected shots that you miss you will be paid 0,00 Euros. Thus you can earn between 0,00 and 60,00 Euros for your shooting. The 10 paid shots were chosen by a random number generator this morning and are in this sealed envelope which I will leave here and which we will open together before calculating your payoffs.

Independent of how many shots you make, you will receive 5,00 Euros for participating. This means that in total you can be paid up to 65,00 Euros (60,00 Euros for your shots + 5,00 Euros for participating).

Once you finish your 300 shots you and I will calculate your payoffs. We will do this by first opening (together) your envelope with the 10 randomly selected shot numbers, then seeing which of the corresponding shots you made and missed, with a 6,00 Euro payment for each make. Then I will pay you your money and you will be free to leave.

Communication is Prohibited:

While you are shooting please do not directly communicate with the rebounder or with me in any way.

Summary:

You will now shoot 300 times with your toes just behind the line marked with tape. Once you have finished you and I will calculate your payoffs. Then I will pay you.

Do you have any questions? If so, please ask now because once you start shooting I will not be able to answer any of your questions.

Thank you for your attention.

You are now free to start shooting your shots. I will announce to you when you have completed your 50th, 100th, 150th, 200th, 250th, and 300th shot so you know how many shots remain.

Rebounding

INSTRUCTIONS: REBOUNDER

You will be asked to rebound each of the 300 shots performed by the shooter. You have a line marked with tape from where you will always pass the ball to the shooter, while facing the shooter squarely. I ask that you always deliver the same pass, a two-handed bounce pass originating from the head. I ask that you try to be as mechanical and repetitive as possible, so the situation changes as little as possible for the shooter, from shot to shot. Before the shooter's first shot, you will stand on your marked line, facing him squarely, and once you hear a clear audible signal you will deliver him the two-handed bounce pass originating from the head. Once you have passed the ball you rotate 180 degrees so that your back is now facing the shooter. You prepare to recover the rebound from the shooter's shot. Once you see the ball make or miss quickly grab the rebound and come back to the marked line. You will wait there, facing the shooter, until I give you a clear audible signal. When you hear the signal this means that you should deliver the two-handed bounce pass originating from the head, to the shooter. You then immediately rotate 180 degrees to await the next rebound, and so forth. I will announce to you when the shooter has completed his 50th, 100th, 150th, 200th, 250th, and 300th shot so you know how many shots remain.

Finally, please avoid any type of direct communication with the shooter, as any such communication can corrupt the scientific validity of the study.

Do you have any questions? If so, please ask now because once the experiment has started I will not be able to answer any of your questions.

Thank you for your attention.

B Appendix: Statistical tests

B.1 Distributions

Claim: The hit streak frequency statistic (H_F) has a normal asymptotic distribution (as $|S| \to \infty$).

Proof: It is sufficient to show that $|S_H| = |\{s \in S : x_{s-1} = x_{s-2} = x_{s-3} = 1\}|$ has a normal asymptotic distribution. Let n be the number of shots, n_1 be the number of hits, r_{1j} be the number runs of hits of length exactly j, and s_{1k} be the number of runs of hits of length k or more.[97] Clearly, the number of (overlapping) sequences of 3 hits in a row satisfies $|S_H| = \sum_{j=3}^{n} r_{1j}(j-2)$, which can easily be simplified to $n_1 - 2r_{12} - r_{11} - 2s_{13}$. Theorem 1 of Mood (1940) shows that for n_1/n fixed, (r_{11}, r_{12}, s_{13}) has a (multivariate) normal asymptotic distribution, and therefore, as a corollary, the asymptotic distribution of $|S_H|$ is normal.

Claim: The probability at each value in the discrete support of the hit streak length statistic (H_L) can by approximated using a normal distribution.

Proof: Let n be the number of shots, n_1 be the number of hits, r_{1j} be the number runs of hits of length exactly j, and s_{1k} be the number of runs of hits of length k or more. The support of H_L is $\{1, 2, \dots, n_1\}$, and its discrete distribution is given by:

$$\mathbb{P}(H_L = \ell) = \begin{cases} \mathbb{P}(s_{12} = 0) & \text{if } \ell = 1 \\ \mathbb{P}(r_{1\ell} \geq 1, s_{1\ell+1} = 0) & \text{if } 1 < \ell < n_1 \\ \mathbb{P}(s_{1n_1} = 1) & \text{if } \ell = n_1 \end{cases}$$

Theorem 1 of Mood (1940) shows that for n_1/n fixed, for $1 \leq \ell < n_1$, $(r_{11}, r_{12}, \dots, r_{1\ell}, s_{1\ell+1})$ has an (multivariate) normal asymptotic distribution. Therefore each probability can be approximated from the associated multivariate normal distribution (with a continuity correction).

Claim: The first order serial correlation statistic and the runs statistic yield the same test for a shooter with a 50 percent hit rate.

Proof: The first order serial correlation statistic is defined as $\rho := \sum_{s=1}^{|S|-1}(x_s - \bar{x})(x_{s+1} -$

[97] Note that $s_{1k} = \sum_{j=k}^{n} r_{1j}$.

$\bar{x})/\sum_{s=1}^{|S|}(x_s - \bar{x})^2$, and the runs statistic can be represented as the number of switches, plus one: $R = 1 + \sum_{s=1}^{|S|-1}[1 - x_s x_{s+1} - (1 - x_s)(1 - x_{s+1})]$. It is easy to show that $R = (1 - \rho)|S|/2 + 1/2$ when the hit rate is fixed at .5, which implies that across all permutations, ρ and R are perfectly correlated; therefore, a runs test will reject the null if and only if the serial correlation test rejects the null.

In the case that shooting performance instead deviates from the design target of 50 percent, the correlation between the statistics is still nearly perfect because $\tilde{R} := -2(|S| - 1)\hat{\sigma}^2\rho + (|S| + 1)/2$, is a close approximation to R, and is perfectly correlated with ρ—the absolute difference satisfies $|R - \tilde{R}| = |2\hat{\mu} - 1|(x_1 + x_{|S|})$, where $\hat{\mu}$ is the fraction of hit shots and $\hat{\sigma}^2$ is the standard deviation, both fixed across permutations.

B.2 Power & Effect Size

Below we report the power of the hit streak statistics (see Section 3) to detect departures from the null hypothesis of consistent shooting (at the $\alpha = .05$ significance level). Further, we consider two tests commonly used in the previous literature: the runs test and the conditional probability test (which compares the hit rate immediately following three consecutive hits to the hit rate immediately following three consecutive misses).

Of the statistics we propose, H_F and H_M outperform the runs and conditional probability tests across all alternative models of hot hand shooting for the parameters we select, and the difference in power is particularly large in the case that a shooter has taken 300 or fewer shots. While all statistics are underpowered at detecting the hot hand when a shooter has taken 100 shots (regardless of the model and chosen parameters), with 300 shots, H_F or H_M are typically adequately powered (though this is only true for certain models of hot hand shooting). These results illustrate the importance of having a particularly large sample size when testing for the existence of the hot hand on the individual level.

We can estimate the magnitude of the hot hand effect by comparing the hit streak momentum statistic H_M to performance when the shooter has not just hit three or more consecutive shots:

$$\text{Magnitude of hot hand effect } = H_M - \frac{\sum_{s \in S_H} x_s}{|S_H^C|}$$

As discussed below, most tests dramatically underestimate the true magnitude of the hot hand effect, as modeled.

Below we detail these results for the three alternative models of the data generating process that we consider: the regime shift model, the positive feedback model, and the hit streak model.

Regime Shift Model

We consider a player with a baseline state ($Y = b$), with hit probability p, and a hot state ($Y = h$) with hit probability $p + \delta$, where $\delta > 0$. The transition probability $q_{ij} := \mathbb{P}(Y = j|Y = i)$, where $i, j \in \{b, h\}$, satisfies $0.75 < q_{hh} < q_{bb}$, with q_{bb} large (near 1). Let π be the stationary distribution, i.e. the fraction of the time in the baseline state. In our simulations we vary $p \in \{.4, .5, .6\}$, $\delta \in \{.1, .2, .3, .4\}$, $q_{bb} \in \{.97, .98, .99\}$, and $\pi \in \{.80, .85, .90\}$, and find that even with 900 shots, all statistics have low power on average, i.e. below .5. If we restrict our parameters to more extreme values $p \in \{.5, .6\}$, $\delta \in \{.3, .4\}$, $q_{bb} \in \{.99\}$, and $\pi \in \{.80, .85\}$, for 900 shots, H_F, H_M and H_L have an average power of .73, .80 and .76 respectively, whereas the runs tests and conditional probability tests have an average power of .6. In these most extreme cases, in which the hot hand increases shooting percentage by 30 or 40 percentage points, the estimated magnitude of the hot hand effect is severely downward biased, with only 9 and 17 percentage point increases in shooting percentage, respectively (for a true effect of 20 the estimated effect is only 4).

Positive Feedback Models

We consider two models of positive feedback and vary the strength of the serial dependence that each employs. Let p be a player's hit rate.

For the first feedback model, which we term the "feedback streak" model, the probability of a successful shot is defined as follows:

$$\mathbb{P}(x_s = 1) = \begin{cases} p + \delta & \text{if } x_{s-1} = 1, x_{s-2} = 1, x_{s-3} = 1 \\ p & \text{otherwise} \end{cases}$$

In our simulations we vary $p \in \{.4, .5, .6\}$ and $\delta \in \{.1, .2, .3, .4\}$. Unsurprisingly, the Runs statistic is poor at detecting the hot hand in this case: the average power is .48 for 300 shots, and .67 for 900 shots. On the other hand, the H_F and H_M statistics are more powerful at detecting these departures from hot hand shooting (.6 and .75, respectively, for 300 shots, and .8 and .9, respectively, for 900 shots), and also more powerful than the conditional probability test (.57 for 300 shots, .8 for 900 shots).[98] In this model the estimated magnitude of the true hot hand effect is not biased downward, as a streak of three or more hits is precisely the cause of the increase in a player's hit rate.

In the second feedback model, which we term the "feedback ladder" model, the probability of a successful shot is defined as follows:

$$\mathbb{P}(x_s = 1) = \begin{cases} p & \text{if } x_{s-1} \neq 1 \\ p + k\delta & \text{if } \exists k < K : x_{s-1} = 1, \ldots, x_{s-k} = 1 \ \&\ x_{s-k-1} \neq 1 \\ p + K\delta & \text{if } \exists k \geq K : x_{s-1} = 1, \ldots, x_{s-k} = 1 \ \&\ x_{s-k-1} \neq 1 \end{cases}$$

with $k, K \in \{1, 2, \ldots, n\}$, where n is the number of shots.

In our simulations we vary $p \in \{.4, .5, .6\}$, $(K, \delta) \in \{3\} \times \{.05, .10, .15\} \cup \{5\} \times \{.02, .03, .04, .05\}$. For $K = 3$, and 300 shots, unsurprisingly, given the first order serial dependence, the runs test performs well with a power of .8. In comparison, the H_F and H_M statistics each have an average power that is as large or larger (.85 and .80, respectively) with H_F outperforming the runs test by up to 10 percentage points for certain parameter values (H_L has a power of .5). The conditional probability test, on the other hand, has a power of .61. All tests are highly powered with 900 shots (except the one based on H_L). The estimated magnitude of the hot hand effect is downward biased, reflecting only 2/3 to 4/5 of the true effect size in this case. For $K = 5$ the results are similar, except H_F outperforms the runs test by 10 percentage points on average (and H_M outperforms it by 5 percentage points). In this case, the estimated magnitude of the hot hand effect has a slight downward bias (9/10) with respect to the true effect size.

Hit Streak Model

We consider a player with a baseline state ($Y = b$), with hit probability p, and a hot state ($Y = h$) with hit probability p_h. If the player is in a baseline state, the player enters into a hot state with probability q, and remains there for exactly K shots, at which point he returns to the baseline state, with the possibility of entering into the hot state again. In our simulations we vary $p \in \{.4, .5, .6\}$,

[98] H_L has a power of .6 for 900 shots.

$p_h \in \{.9, .95, 1\}$, $q \in \{.01, .02, .03\}$, and $K \in \{10, 15, 20\}$. For 300 shots, the runs and conditional probability tests have an average power of .62 and .64, respectively, whereas H_F, H_M and H_L have an average power of .75, .82, and .70, respectively. For 900 shots, the runs and conditional probability tests have an average power of .86 and .84, whereas H_F, H_M and H_L have an average power of .93, .96, and .76, respectively.

Table 3: *Linear probability model hit rate over for the player RC and JNI player 6 (with fixed session effects, permutation p-values equal proportion of permuted (session strata) data where coefficient exceeds the realized coefficient (one-sided))*

	The player RC		JNI6	
	Main	Categories	Main	Catagories
constant	.609	.678	.585	.706
	(.991)	(.996)	(.971)	(1.000)
Hit 3+	.069***		.120***	
	(.005)		(.001)	
Hit 2		−.109***		−.068
		(.004)		(.113)
Hit 1		−.055**		−.125***
		(.048)		(.007)
Missed 1		−.046*		−.160***
		(.083)		(.002)
Missed 2		−.098**		−.121**
		(.021)		(.036)
Missed 3+		−.078*		−.082*
		(.059)		(.068)

50,500 Permutations (session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided)

C Appendix: Supplementary tables and figures

C.1 Supplementary Analysis

In Section 4.2 we mention that the results of a two-sample proportion test (see Figure 2 for graph) could be driven by selection bias at the session level. To control for this possibility we estimate the marginal effect of having just completed a run of three or more hits on RCs probability of hitting the next shot using a linear probability model with session fixed effects. Under the null hypothesis, the indicator variable for hitting the previous three or more shots is a treatment that is assigned at random to the player. In the first column of Table 3, we present the coefficient corresponding to the marginal effect of hitting three or more shots in a row (*Hit* 3+) on the probability of hitting the next shot (p-value in parenthesis).[99] These marginal effects clearly corroborate the estimates

[99]The p-values are computed from the distribution of the coefficients under the null, where the approximated distribution is generated to arbitrary precision via a session-level permutation of shots. We also estimated the p-value for the model using asymptotic assumptions and robust standard errors (if there are session level effects on a player's hit rate, errors will be heteroskedastic if shots are iid Bernoulli at the session level). The asymptotic p-values computed using robust standard errors were much lower then the approximation of the exact p-values under the null reported

presented in Figure 2, as well as the associated two-sample proportion test on their differences.[100] In column two of Table 3 we estimate the coefficients of a fixed-effects linear probability model with indicators variables corresponding to the five mutually exclusive shooting situations in the right panel of Figure 2 (*Hit* 2, *Hit* 1, *Miss* 1, *Miss* 2, *Miss* 3+).[101] When controlling for these session level effects, the results of the proportion test are corroborated: RC still has a significantly lower shooting percentage in all five of these shooting situations, with the significant coefficients on *Hit* 2 and *Hit* 1, suggesting that this is a hot hand effect and not a cold hand effect. These effect size estimates are biased downward from the true effect for two reasons: (1) with player fixed (mean) effects, when *Hit* 3+ = 1 the player has performed at above mean levels in the previous three shots from the same finite dataset that the mean is calculated for, and therefore the effect size will have a downward bias, (2) not all shots taken when *Hit* 3+ = 1 are taken in a hot state, and therefore the true difference between a hot state and cold state is underestimated.

C.2 Supplementary Tables & Figures

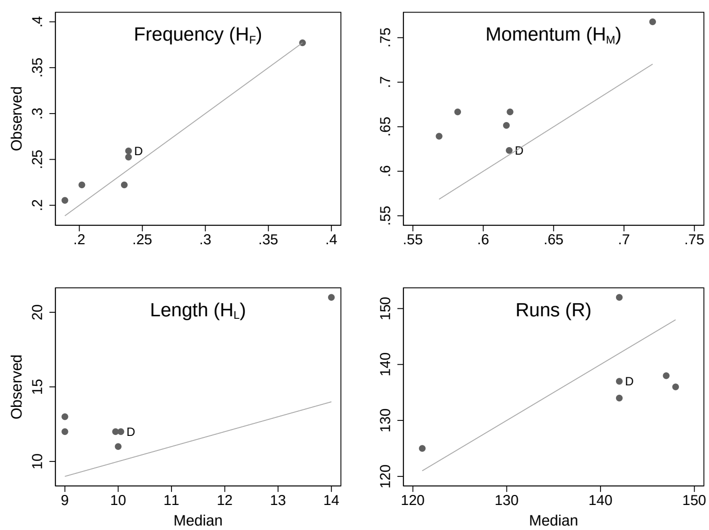


Figure 8: *Observed vs. median hit streak statistics for the player RC, where median is based on the exchangeability assumption (the session labeled was a doublesession)*

---

here, but with only 5 clusters, there is a potential for robust standard errors to be severely biased (Angrist and Pischke (2008), pp. 221-240).

[100] The marginal effects and their significance do not differ substantively under a logit model (which should be expected, Angrist and Pischke (2008), p. 103).

[101] Hitting three or more shots in a row serves as the base category.

Table 4: *For each shooter in the three session panel, and in each phase (and overall), the hit percentage is reported after having hit 3 or more in row and after all other recent shot histories. The number of shots is reported in brackets.*

	Phase 1		Phase 2		Overall	
Shooter	Hit 3 or more	other	Hit 3 or more	other	Hit 3 or more	other
1†	63.9	55.5	66.7	58.5	65.8	57.5
	[61]	[236]	[141]	[453]	[202]	[689]
2	55.2	47.0	55.7	50.4	55.6	49.2
	[29]	[268]	[88]	[506]	[117]	[774]
3	50.0	49.2	59.2	58.0	57.1	54.9
	[38]	[258]	[125]	[469]	[163]	[727]
4	40.9	41.1	41.8	44.2	41.6	43.1
	[22]	[275]	[55]	[539]	[77]	[814]
5	55.6	44.4	47.7	47.4	50.5	46.5
	[36]	[261]	[65]	[529]	[101]	[790]
6	40.0	31.5	28.9	42.4	31.3	38.7
	[10]	[286]	[38]	[556]	[48]	[842]
7	59.5	51.8	61.0	52.2	60.6	52.1
	[42]	[255]	[100]	[494]	[142]	[749]
8	51.1	56.3	59.1	53.9	56.9	54.7
	[45]	[252]	[115]	[479]	[160]	[731]

†Shooter 1,“RC”, had two additional identically conducted sessions. In these sessions his hit percentage was 72.5 [178] and 65.4 [416] respectively.

Table 5: *Linear probability model of the Panel's hit rate over in each Phase (with fixed session effects, permutation p-values equal proportion of permuted (session strata) data where coefficient exceeds the realized coefficient (one-sided))*

	Phase 1		Phase 2		Overall	
	Main	All	Main	All	Main	All
constant	.471	.519	.510	.541	.497	.533
	(.9142)	(.9590)	(.9283)	(.9681)	(.9753)	(.9942)
Hit 3+	.048**		.031**		.035***	
	(.0441)		(.0407)		(.0086)	
Hit 2		−.058*		−.021		−.032*
		(.0696)		(.1825)		(.0593)
Hit 1		−.053*		.001		−.017
		(.0533)		(.4287)		(.1370)
Missed 1		−.017		−.051***		−.038**
		(.2700)		(.0100)		(.0139)
Missed 2		−.083**		−.030*		−.047**
		(.0152)		(.0940)		(.0100)
Missed 3+		−.053*		−.064***		−.059***
		(.0535)		(.0034)		(.0013)

p-values in parentheses

50,500 Permutations (session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided, right)

Table 6: *Linear probability model of hot streak performance (with fixed session effects, permutation p-values equal proportion of permuted (session strata) data where coefficient exceeds the realized coefficient (one-sided))*

	Main Effect					Categories				
	RC	JNI6	Panel	GVT	JNI	RC	JNI6	Panel	GVT	JNI
constant	.609	.585	.497	.478	.622***	.678	.706	.533	.529	.587
	(.991)	(.971)	(.975)	(.864)	(.004)	(.996)	(1.000)	(.994)	(.991)	(.621)
Hit 3+	.069***	.120***	.035***	.050**	−.036					
	(.005)	(.001)	(.009)	(.014)	(.674)					
Hit 2						−.109***	−.068	−.032*	−.021	.044
						(.004)	(.113)	(.059)	(.213)	(.815)
Hit 1						−.055**	−.125***	−.017	−.069***	.071
						(.048)	(.007)	(.137)	(.009)	(.963)
Missed 1						−.046*	−.160***	−.038**	−.052**	.012
						(.083)	(.002)	(.014)	(.029)	(.303)
Missed 2						−.098**	−.121**	−.047**	−.008	.043
						(.021)	(.036)	(.010)	(.244)	(.587)
Missed 3+						−.078*	−.082*	−.059***	−.079***	−.031**
						(.059)	(.068)	(.001)	(.002)	(.014)

p-values in parentheses

50,000 Permutations (session strata)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ (one-sided, right)