Psychology of Sport and Exercise 14 (2013) 21–27

Contents lists available at SciVerse ScienceDirect

Psychology of Sport and Exercise

journal homepage: www.elsevier.com/locate/psychsport

Review

The "hot hand" reconsidered: A meta-analytic approach

Simcha Avugos [a,d,*], Jörn Köppen [b], Uwe Czienskowski [c], Markus Raab [b], Michael Bar-Eli [a,d]

[a] Department of Business Administration, Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[b] Institute of Psychology, German Sport University, Cologne, Germany
[c] Max Planck Institute for Human Development, Berlin, Germany
[d] Zinman College of Physical Education and Sport Sciences, Wingate Institute, Israel

ARTICLE INFO

Article history:
Received 31 August 2011
Received in revised form
10 July 2012
Accepted 16 July 2012
Available online 25 July 2012

Keywords:
Streaks
Momentum
Sport
Effect size
Moderator analysis

ABSTRACT

Objectives: Extensive searches for the "hot hand" have been performed in a variety of sports since the pioneering study of Gilovich, Vallone, and Tversky (1985), but empirical evidence for the existence of the effect is still fairly limited. The current review reconsiders the hot hand in sports using a meta-analytic approach.
Design: Mean effect size and 95% confidence interval were determined using a random effects model. Heterogeneity of the mean effect size was examined applying Cochran's Q test and the "75 percent rule".
Method: To be included in the meta-analysis, studies had to provide an empirical investigation of the hot hand phenomenon related to sport and exercise behavior. Approximately 250 papers were located, but the final dataset included only 22 publications that met inclusion criteria, with 30 studies and 56 independent effect sizes. The articles extended over a period of twenty-seven years from 1985 until 2012.
Results: The analysis of the effects yielded a minor positive mean effect size of .02, $p = .49$, using a random effects model, which is sufficient evidence for arguing against the existence of the hot hand. Due to the limited sample of studies available, only a few candidate-variables could be extracted and further examined as potential moderator variables. However, none of the considered variables had the power to explain the heterogeneity of effect sizes.
Conclusions: The present study provides additional support for Gilovich et al.'s claim that a general hot hand effect probably does not exist in sport. The scientific implications of this review for prospect advances in the field are presented and discussed.

We all experience daily the unavoidable need to make decisions. Some are of minor consequence (e.g., whether to take an umbrella on a cloudy day), while others may be critical (e.g., whether to undergo a risky medical treatment). Many decisions are based on beliefs regarding the probability of uncertain events. In the workplace, for example, managers and employees must make decisions concerning the future profit of a company, the demand for a commodity, the market performance of a new product, or the expected job performance of an applicant. In sports, the choices that coaches and athletes make in response to the environment or competitor can make a vital difference to the results. In the game of basketball, for example, passing the ball to the team's best shooter, who is closely guarded by the opposing team, might be a non-optimal decision if other players have a better chance of scoring.

In their first joint paper, which discussed belief in the law of small numbers, Tversky and Kahneman (1971) suggested that both laypersons and knowledgeable respondents often err in their judgment of random sampling. Apparently, people's intuitive predictions and judgments of probability do not always comply with simple probabilistic laws or the rules of statistics, mainly when considering rich contexts. It has been argued that statistical intuitions develop from our experience with randomizing devices such as coin tosses, throws of the dice, or spins of the roulette wheel (Lagnado & Sloman, 2004). Tversky and Kahneman (1974) described three heuristics that are commonly used when assessing probabilities and predicting values: representativeness, availability, and adjustment from an anchor. These rules of thumb are highly economical and quite useful, but sometimes lead to systematic and predictable errors. According to the representativeness heuristic, inferences are made with no consideration of either prior probabilities or sample sizes, and thus may result in fallacious judgments.

Rational judges are expected to make probability judgments in accordance with their knowledge about the subject matter, the

* Corresponding author. Department of Business Administration, Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev, P.O. Box 653, Ben-Gurion Blvd., 84105 Beer-Sheva, Israel.
E-mail address: avugos@bgu.ac.il (S. Avugos).

1469-0292/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.psychsport.2012.07.005

laws of probability, and their own judgmental heuristics and biases (Tversky & Kahneman, 1974). Therefore, it has been suggested that an awareness of the biases to which these heuristics lead can improve judgments and decisions in situations of uncertainty encountered in the normal course of life (see the review of Oskarsson, Van Boven, McClelland, and Hastie (2009) for an overview of judgment processes in various domains, such as lotteries, sports performances, stock prices, and others).

Misperception of chance events and a tendency to neglect sample sizes were both demonstrated in the "hot hand" studies initiated by Gilovich and his colleagues (Gilovich et al., 1985; see also Tversky & Gilovich 1989a, 1989b). The following section describes the hot hand phenomenon from the world of sport, where it is attributed to fallacious judgment associated with the belief in the law of small numbers and the representativeness heuristic.

Hot hand research

The hot hand phenomenon is known to everyone who plays or watches the game of basketball: Players and fans believe that after a run of successful baskets a player is more likely to succeed with the next shot as well. The hot hand belief is usually justified by causal explanations. For example, a "hot" player is likely to be highly motivated and more confident in his or her ability, and thus the probability of success in this player's further attempts is expected to increase. Athletes believe in the hot hand because they themselves have experienced it.

Gilovich and his associates (1985) described the widespread belief in the hot hand using statistical terms, which were then empirically tested in the game of basketball. Several within-game measures (e.g., analyses of conditional probabilities, run counts, and serial correlations), and one between-game measure (a comparison between the observed and the expected variability in shooting percentages across games using the Lexis ratio), were applied. Three sets of real data were used to examine the existence of the effect: field goal records, pairs of free throw shots of professional basketball players over two NBA seasons, and data from a controlled shooting experiment with varsity basketball players. However, no evidence for the hot hand was found in any of these data sets.

When considering the field goal data, contrary to the hot hand expectations, (a) players' probabilities of making a shot were usually somewhat lower after having made successful shots than after having missed shots, (b) the average serial correlation between the outcome of successive shots was nearly zero, (c) no significant difference was found between the observed and expected number of runs of hits and misses in the data for each player, across all games and within individual games, and (d) the "hot" or "cold" spells in the players' hit rate did not occur more frequently than would be expected by chance (i.e., a test of stationarity). Shots were also independent during free throw shooting during a game of the Boston Celtics and in an experiment with Cornell's varsity teams (Gilovich et al., 1985). Both data sets were uncontaminated by external factors such as players' shot selection or defensive pressure (the changes in shooting accuracy have been recently discussed in relation to such activities by Sun and Wang (2012)). This was also evident when considering only pairs of shots taken in close proximity (Tversky & Gilovich, 1989b). The overall results indicate that the chances of making a given shot are largely independent of prior performance, however the outcomes of previous shots may affect a player's willingness to take the next shot (Tversky & Gilovich, 1989a).

Gilovich and his associates (1985) suggested that the consistent belief in the hot hand is merely an illusion, due to either a memory bias or a cognitive bias associated with the representativeness heuristic. According to the memory bias, streaks of hits (e.g., HHHHHH) are more memorable than hit-miss alternations (e.g., HMMHMH), and therefore the observer is likely to overestimate the frequency of streaks. Otherwise, according to the cognitive bias line of reasoning, people usually expect random sequences to contain more alternations and fewer long runs than actually occur, and thus tend to detect streaks even when none exist. At this stage, this rationale was exclusively related to basketball shooting (Gilovich, 1991, pp. 9–21).

The majority of follow-up studies were also unable to detect the hot hand, either in professional basketball (Adams, 1992; Koehler & Conley, 2003; Vergin, 2000), or in other sports including baseball (Albert, 2008; Albert & Bennett, 2001; Albright, 1993; Frohlich, 1994; Siwoff, Hirdt, & Hirdt, 1988; Vergin, 2000), golf (Clark, 2003a, 2003b, 2005), and dart throwing (Gilden & Wilson, 1995). Some fragile support for the hot hand, however, was found in few motor tasks that can be quickly and regularly repeated (all involve relatively simple arm motions), namely horseshoe pitching (Smith, 2003), bowling (Dorsey-Palmateer & Smith, 2004; Frame, Hughson, & Leach, 2004), and tennis (Klaassen & Magnus, 2001). All in all, a survey of twenty years of research on the topic (Bar-Eli, Avugos, & Raab, 2006) showed that empirical evidence for the effect is rather limited, and demonstrations of hot hands per se are rare and often weakly supported.

Although almost the implicit statistics suggests that unusual streaks are singular events (associated with great talent combined with pure luck), still some researchers strongly believe that streaks often occur. Since the effect sizes are likely to be small, the standard statistical tests are assumed to have not enough power to detect them. Due to the small effect size, it has been suggested that a large number of observations are required to detect significant streaks (Dorsey-Palmateer & Smith, 2004; Miyoshi, 2000; Swartz, 1990; Wardrop, 1999). From a statistical point of view, such an argument is hardly accepted. Although some statistical tests are more sensitive than the others, the weakness of a specific test should be related to not having enough data to detect a small effect with sufficient sensitivity and accuracy. Obviously, either the theory is just a poor one, or the predicted effect simply does not exist. In any case, this line of reasoning has no power to explain why players and fans believe in the hot hand if they are totally unable to detect it.

In medicine, for example, even a very small effect is valuable if a specific treatment might save the life of few patients. In sports, under very specific circumstances, the belief in the hot hand might serve as an adaptive strategy for allocation decisions, even if shots really are independent events (Burns, 2004; Raab, Gula, & Gigerenzer, 2012). However, if the effect size is too small to be detected, the scientific as well as practical implications of such an effect are questionable. In any case, the hot hand effect is probably much smaller than athletes and fans believe it to be.

To summarize, the outcomes of hot hand research vary from seeming to support to seeming to argue against the assumed existence of the effect under study. When reviews of research are presented and described only qualitatively (e.g., Bar-Eli et al., 2006), the results of conflicting studies can be confusing. It also might be too tempting for authors of narrative reviews consciously or unconsciously to select and describe studies to support their own understanding of the literature, as pointed out by Rosenthal and DiMatteo (2001). In the case of the hot hand, meta-analysis is a powerful method to use for such position determination, applying a pure vote counting procedure. It allows us to arrive at conclusions that most likely are more accurate and more credible than can be presented in a narrative review.

Thus, the purpose of the current study was twofold. The first goal was to objectively address the question of whether streaky performance (i.e., the tendency for outcome sequences to form streaks) really occurs in sport activities. The results of such an assessment are of interest on various accounts, in particular for sport psychologists, since the issue relates directly to the efficacy of intentional behavior. In light of the intuitive strong belief in the hot hand, it would be instructive to clarify the current state of research using a meta-analytic tool. Combining evidence across studies would also allow us to validate the existence of the effect across various sporting events, beyond basketball shooting.

The second goal was to determine whether any personal or situational factors induce the existence of the phenomenon, as suggested, for example, by Oskarsson et al. (2009). Specifically, we examined the potential causal influence of three moderator variables: participants (players and teams), time interval of assessment (within-game and across-game streaks), and type of sport (individually-performed sports and team sports). Correlations between these variables and effect sizes are very helpful for advancing future research, as well as for promoting practical applications in various domains, including sport. Based on all available data (as well as the 2006 narrative literature review of Bar-Eli et al.), we anticipated that our review would provide additional evidence against a weakly confirmed empirical effect.

Method

Literature search

We used three procedures to locate studies for our analysis. First, an extensive search was conducted through the *"Web of Science"* citation index. This database provided a list of 254 publications referencing Gilovich et al.'s (1985) hot hand study, of which only 38 papers met our inclusion criteria. An additional search for relevant studies was conducted within the most important online databases for education, psychology, and sports (*PsycInfo*, *ERIC*, *SportDiscus*, and *Dissertation Abstracts International*). We used "hot hand", "streaks", "flow", and "momentum" as the keywords, as well as modifications of these keywords, such as "streaky". Studies initiated before 1985 that either did not address the current problem domain or provided no compatible statistics, were excluded (e.g., Adler & Adler, 1978; Hoffman, 1983; Iso-Ahola & Mobily, 1980). Alan Reifman's hot hand website (http://thehothand.blogspot.com) and recent book (Reifman, 2012), which summarizes 25 years of psychological and statistical research on sports streaky behavior, were also screened for additional data.

Not every computer-assisted search will be complete. Therefore, the reference lists of the studies and the Table of Contents of several journals that had already published hot hand articles were manually used to cross check and extend our search (*The American Statistician, Applied Statistics, Chance, Cognitive Psychology, International Journal of Sport Psychology, Journal of Sport & Exercise Psychology, Journal of Sport Behavior, Journal of the American Statistical Association, Perceptual and Motor Skills, Psychonomic Bulletin & Review, The Sport Psychologist, The Statistician*). These procedures provided 13 additional relevant studies that were not located by the *"Web of Science"*.

A search for so-called "grey" publications such as university reports and unpublished dissertations, using national and international e-mail lists in sport science (e.g., German Sport Science e-mail list of sport scientists), provided one additional study (Frame et al., 2004) to be meta-analyzed. Eventually, we attempted to extract information from a total of 52 studies that appeared to meet our inclusion criteria.

Inclusion and exclusion criteria

To be included in the meta-analysis, studies had to provide an empirical investigation of the hot hand phenomenon related to sport and exercise behavior. Sports and exercise behavior is defined as sports that are organized by and associated to Olympic sport associations. Thus, activities such as chess or horseshoe pitching were excluded. After a preliminary review, we eliminated a substantial number of studies, over several hundred publications, that were either conducted in other academic fields outside the sport domain (e.g., hot hands among mutual fund markets: Cheng, Pi, & Wort, 1999; the hot hand in enduring political rivalries: Gartzke & Simon, 1999), or that discussed a related sport topics (rather than genuine sport activities as above defined. e.g., the hot hand in sport betting market: Dixon & Pope, 2004; the hot hand in the domain of roulette: Croson & Sundali, 2005). Finally, we checked whether typical binary outcomes of hit/miss within-game or win/loss between games are used for effect sizes.

We rejected several studies out of the remaining 52 relevant sport publications for the following reasons: (a) the study provided insufficient information to compute an effect size (missing effect or sample sizes; e.g., Gould, 1989), (b) the study represented a theoretical (as opposed to empirical) analysis (e.g., Albert & Williamson, 2001; Korb & Stillwell, 2003), (c) the study discussed the belief in the hot hand, but used no actual performance data to investigate it (e.g., Burns & Corpus, 2004; Hales, 1999; Hooke, 1989), (d) the study reported faulty results, due to methodological errors in replicating Gilovich et al.'s (1985) design (Hancock & Frisch, 2005), (e) the study reported faulty results related to the coding of the raw data (Larkey, Smith, & Kadane, 1989), or (e) the study merely commented on previous research reports (e.g., Gula & Raab, 2004). Final inclusion and exclusion decisions were based on an existing coding scheme, which will be discussed in the following section.

Overall, our dataset included 22 publications that met inclusion criteria, with 30 studies and 56 independent effect sizes. The articles extended over a period of twenty-seven years, starting with Gilovich et al.'s, (1985) research and ending with the study by Avugos, Bar-Eli, Ritov, and Sher (in preparation).

Coding

After the studies were identified, we developed a coding scheme. The coding scheme included 26 items of both general information (e.g., authorship, year of publication, sampling methods, experimental design) and variables of interest within each study (e.g., type of sport, expertise of participants, group size, statistical data). The variables of interest were developed based on arguments embedded in the literature, suggesting that the hot hand might be moderated by the type of sport or required movement (e.g., Reifman, 2012), the expertise of athletes (e.g., Avugos et al., in preparation; Köppen & Raab, 2012), or the sensitivity of measures reflected by sample size and type of statistical data (e.g., Tyszka, Zielonka, Dacey, & Sawicki, 2008; Wardrop, 1995). Moderators were labeled as hot hand in the individual player's level vs. team level, within-game vs. across-game streaks, and type of sport as individually-performed sports vs. team sports. Two coding assistants independently coded the studies, and disagreements (fewer than 10 instances) were resolved by discussion. In the final stage, a statistical consultant re-examined the codings and all combinations of variables.

Computation and analysis

The studies used for our analysis differed considerably in their methodological approaches, as well as in their statistical information. Meta-analytic integrations usually require the transformation of all statistical measures into a common effect size index, which allows researchers to compare different effects and finally to compute a mean effect size across all studies. This averaged effect would estimate to some degree the “true” population effect size, which is assumed to be reflected in the results of each individual study.

To make the direct comparisons possible for our analysis, we extracted four different values (correlation, t, z, and χ^2 value) and converted them to a common metric value (point-biserial correlation), “r”. These transformations were based on each player's sequential performance record, indicating either a stable (a positive correlation coefficient) or a variable (a negative correlation coefficient) performance. We used the Comprehensive Meta-Analysis Software (Borenstein, Hedges, Higgins, & Rothstein, 2005) for our calculations.

The relative weight of each study depended on the number of effect sizes reported: the higher the number of effects reported, the higher the study's weight in our analysis. We classified the effect sizes into one of the following three categories: (a) “positive” – if the study supported the existence of the hot hand, (b) “negative” – if the study provided evidence against the existence of the hot hand, and (c) “not specified” – if the study reported inconclusive results or if the effect size was not mentioned. Overall, we identified and coded 32 positive effects representing a stable performance and 24 negative effects representing a variable performance.

Based on the methods described by Lipsey and Wilson (2001), we calculated the mean effect size and 95% confidence interval using a random effects model. Heterogeneity of the mean effect size was examined using Cochran's Q test (Hedges & Olkin, 1985: The Q value is calculated as the weighted sum of squared differences between individual study effects and the pooled effects across studies). In addition, we used the “75 percent rule” (Hunter & Schmidt, 2004) to determine the percentage of total variation across the studies due to heterogeneity. If the observed variance percentage due sampling error was less than 75 percent, which indicates heterogeneity of effects, we performed moderator analyses.

Results

Overall effect size

The effect sizes are presented in Table 1. Column 2 shows the transformed correlation values: a positive result supports the existence of the hot hand, while a negative result provides evidence against the existence of the effect. The higher ($0 < r < 1$) or lower ($-1 < r < 0$) the value, the stronger the position; a value of zero implies a non-conclusive result. The correlation values vary considerably (ranging from -1 to .92). Overall, 56 effect sizes were calculated from the 22 hot hand articles. The analysis of the effects yielded a small and positive mean effect size of .02 (95% CI = $-.03$, .06), using a random effects model. This is sufficient evidence for arguing against the existence of the hot hand. The mean effect size was heterogeneous ($Q = 377.02$, $df = 55$, $p < .01$), and the observed variance percentage due sampling error, var(rc), was less than 75%, which supported the examination of moderating variables.

Analysis of moderating effects

Several candidate-moderating variables that might influence the existence of the hot hand were considered. Among them were personal (e.g., skill, level of expertise) and situational factors (e.g., the structure of different sport situations. Adams, 1992; Smith, 2003). More explicitly, Hastie and Dawes (2010) suggested that in nonreactive, uniform-playing-field sports, subtle sequential dependencies would manifest themselves in performance, while in chaotic, “in-your-face”, player-on-player reactive sports, such patterns are not to be expected (see pp. 145–146). Reifman (2012) attributed the

Table 1
Statistics for all studies included in the meta-analysis of hot hand research.

Name of study	Correlational point estimate	Original value	Sample size	95% CI	p Value
Gilovich et al., 1985, S2a	−.51	$Z = -.56$	9	−.88 to .24	.17
Gilovich et al., 1985, S2b	−.73	$t = -2.79$	9	−.94 to −.12	.02
Gilovich et al., 1985, S2c	−.76	$t = -3.14$	9	−.95 to −.20	.01
Gilovich et al., 1985, S2d	−.86	$t = -4.42$	9	−.97 to −.45	<.01
Gilovich et al., 1985, S3	.00	$r = .00$	9	−.66 to .66	1.00
Gilovich et al., 1985, S4a	−.02	$r = -.02$	26	−.40 to .37	.94
Gilovich et al., 1985, S4b	−.96	$Z = 2.00$	26	−.98 to −.92	<.01
Tversky & Gilovich, 1989a, 1a	−.04	$r = -.04$	9	−.69 to .64	.92
Tversky & Gilovich, 1989a, 1b	.00	$r = .00$	9	−.66 to .66	1.00
Tversky & Gilovich, 1989b	−.02	$r = -.02$	18	−.48 to .45	.94
Miller & Weinberg, 1991, S2a	−.12	$t = 2.17$	328	−.22 to −.01	.03
Miller & Weinberg, 1991, S2b	−.17	$t = 3.17$	328	−.28 to −.07	<.01
Adams, 1992, a	.15	$\chi^2 = .86$	45	−.15 to .42	.33
Adams, 1992, b	.34	$\chi^2 = 5.17$	45	.05 to .58	.02
Adams, 1992, c	.53	$\chi^2 = 12.78$	45	.28 to .71	<.01
Albright, 1993	−.08	$r = -.08$	40	−.38 to .24	.63
Stern & Morris, 1993	.03	$r = .03$	40	−.28 to .34	.86
Gilden & Wilson, 1995, S1	.40	$t = -2.69$	40	.10 to .63	.01
Gilden & Wilson, 1995, S2	.92	$t = -3.94$	5	.17 to .99	.03
Gilden & Wilson, 1995, S3	−.19	$t = .47$	8	−.79 to .60	.67
Gilden & Wilson, 1995, S4	.79	$t = -1.80$	4	−.72 to 1.00	.29
Wardrop, 1995, a	.15	$Z = .15$	9	−.57 to .74	.71
Wardrop, 1995, b	−1.00	$Z = -4.30$	9	−1.00 to −1.00	<.01
Vergin, 2000, S1	−.24	$Z = -.24$	28	−.56 to .15	.23
Vergin, 2000, S2a	−.04	$Z = .04$	29	−.40 to .33	.84
Vergin, 2000, S2b	−.04	$Z = .04$	29	−.40 to .33	.84
Clark, 2003a, S1	−.50	$r = -.50$	35	−.71 to −.20	<.01
Clark, 2003a, S2	−.32	$r = -.32$	31	−.61 to .04	.08
Clark, 2003b	−.42	$r = -.42$	25	−.70 to −.03	.04
Koehler & Conley, 2003	−.03	$Z = -.03$	23	−.44 to .38	.88
Frame et al., 2004, a	.08	$t = -1.65$	436	−.02 to .17	.10
Frame et al., 2004, b	.12	$t = 2.49$	436	.03 to .21	.01
Clark, 2005	−.01	$r = -.01$	35	−.34 to .32	.95
Dumangane, Rosati, & Volossovitch, 2009	.47	$r = .47$	4	−.90 to .99	<.01
Arkes, 2010, a	.12	n/a	435	.03 to .22	<.01
Arkes, 2010, b	.18	n/a	213	.04 to .30	<.01
Arkes, 2010, c	.19	n/a	113	.00 to .36	<.05
Arkes, 2010, d	.09	n/a	383	−.02 to .18	<.10
Arkes, 2010, e	.10	n/a	373	.00 to .20	<.05
Palacios, 2010	.09	$\chi^2 = 1.58$	162	−.06 to .25	.21
Jordet, Hartman, & Vuijk, 2011	.24	n/a	85	.02 to .43	.03
Yaari & Eisenmann, 2011, a	.13	n/a	848	.06 to .19	<.01
Yaari & Eisenmann, 2011, b	.14	n/a	859	.07 to .21	<.01
Yaari & Eisenmann, 2011, c	.14	n/a	839	.07 to .21	<.01
Yaari & Eisenmann, 2011, d	.07	n/a	826	.00 to .13	.06
Yaari & Eisenmann, 2011, e	.09	n/a	830	.02 to .16	<.01
Livingston, 2012, a	.03	n/a	156	.01 to .05	<.01

Table 1 (*continued*)

Name of study	Correlational point estimate	Original value	Sample size	95% CI	*p* Value
Livingston, 2012, b	.02	n/a	144	.00 to .04	>.10
Livingston, 2012, c	−.02	n/a	77	−.05 to .01	>.10
Livingston, 2012, d	.03	n/a	76	−.01 to .06	>.10
Avugos et al., in preparation, S1a	.00	$r = .00$	26	−.39 to .39	.99
Avugos et al., in preparation, S1b	.00	$r = .00$	26	−.39 to .39	1.00
Avugos et al., in preparation, S2a	−.03	$r = -.03$	32	−.38 to .32	.87
Avugos et al., in preparation, S2b	.00	$r = .00$	32	−.35 to .35	.99
Avugos et al., in preparation, S2c	.04	$r = .04$	32	−.31 to .39	.81
Avugos et al., in preparation, S2d	−.02	$r = -.02$	32	−.36 to .33	.92
Overall (random model)	**.02**	–	–	−.03 to .06	.49

Note. The studies under review are listed in chronological order. Numbers (S1, S2, etc.) were assigned if several studies were included within one article; letters (a, b, c etc.) were added if more than one effect size was reported in the study. CI = confidence interval; *p* value = the probability of obtaining a value at least as extreme as the one that was actually observed, assuming that the null hypothesis is true; n/a = not applicable. The effect size had to be calculated.

existence of the hot hand to the relative simplicity or complexity of the executed arm motion. Reifman pointed out that the hot hand was detected in sports tasks that involve brief, relatively simple quickly repeated arm motions (compared to more elaborate motions, such as throwing a pitch in baseball), due to the mechanism of human motor memory. In the case of basketball shooting, it involves repeated attempts to release the ball from the same location, at the same angle, with the same force, etc, which could barely reproduced exactly again and again. Therefore, it could hardly be expected from a basketball player to necessarily shoot the very same way on two successive throws, even if he or she is doing their best.

Given the limited quality of our dataset, we examined the influence of merely three potential variables of interest: participants, time interval of assessment, and type of sport on the hot hand. The first variable (participants) refers to both individual players (47 effect sizes) and teams (9 effects). Sample sizes vary from 4 to 859. The second variable refers to the time interval of assessment of the hot hand effect, and has two groups: Hot hand streaks detected within a single game would be considered "within-game" effect, whereas hot hand streaks detected across games over the course of a season or several seasons, would be considered "across-game" effect. Overall, we identified and coded 37 across-game effects and 19 within-game effects. The third variable (type of sport) refers to the various sports in which the hot hand phenomenon was examined. Research was conducted in nine different sports: basketball (31 effect sizes), golf (10 effects), baseball (4 effects), billiards (3 effects), volleyball (2 effects), bowling (2 effects), dart throwing (1 effect), handball (1 effect), soccer (1 effect), and golf with darts combined (1 effect). However, since averaging effect sizes across less than five studies could not provide a sufficiently robust estimate of a mean effect size, we grouped the effects for this variable into individually-performed sports (17 effects: golf, billiards, bowling, dart throwing, and golf with darts combined) vs. team sports (39 effects: basketball, baseball, volleyball, handball, and soccer).

The effect sizes and confidence intervals (CI), as well as the Q values for each moderator variable, are presented in Table 2. Recall that the mean effect size was heterogeneously distributed ($Q = 377.02$, $df = 55$, $p < .01$, var(rc) < 75%). However, none of the candidate moderators exhibited a powerful ability to explain the variability across studies. In other words, the mean effect size was not influenced by any of the selected variables. An analysis of the effect sizes of both players and teams indicated that there is still a large random effects variance, due to other uncontrollable influences on the observed effect sizes. The same pattern emerged when both the time interval of assessment and the type of sport were considered.

Since we are unable to rule out alternative explanations of the results patterns (especially due to our limited sample of studies), there is no evidence for any kind of moderating effect of the selected variables. It is even possible that each reported effect was generated by a variety of factors, all different from the hot hand.

Publication bias

Obviously, one could not obtain every piece of information ever generated about the hot hand. In many cases, a journal's decision to publish is influenced by the internal validity level of the study and statistical significance of the results obtained (Rosenthal & DiMatteo, 2001). Therefore, methodologically top quality studies and those with larger effects are more likely to be published (known as the "file drawer problem"). Thus, a publication bias might arise if some unpublished hot hand papers would change the direction or size of the results reported in our analysis. To reduce the potential of such bias, our search strategy included grey literature search as well as associations in sport science community to locate missing data. Alan Reifman's hot hand site on the web and recent book were also searched thoroughly.

To address the possibility of potential publication bias, we used Egger's funnel plot graphical presentation (Egger, Smith, Schneider, & Minder, 1997). Overall effect estimates were plotted against sample size. A symmetrical funnel is observed in the absence of unpublished hidden studies. However, in some cases it is difficult to determine visually whether the funnel is symmetrical or asymmetrical. Therefore, an additional analysis using Egger's regression test was applied (intercept = −1.70, $p < .05$), which confirmed a skewed and asymmetrical funnel, implying the possibility of potential presence of publication bias.

Discussion

The first goal of this study was to provide an objective assessment of the non-random occurrence of streaky performance in sport. Since

Table 2
Effect sizes by moderators in the meta-analysis of hot hand research.

Moderator	k	Np	Mean r	95% CI	Q	df	p Value
Players	47	2455	.02	−.03 to .07	348.65	46	<.01
Teams	9	987	−.02	−.12 to .09	26.80	8	<.01
Within-game effect	19	514	.00	−.16 to .16	147.03	18	<.01
Across-game effect	37	2928	.03	−.01 to .08	228.03	36	<.01
Team sports	39	2325	-.07	−.15 to .01	296.04	38	<.01
Individual sports	17	1117	.04	.00−.08	65.76	16	<.01

Note. K = number of effects; Np = number of participants, based on the dataset of each study; Mean r = averaged correlation value; CI = confidence interval; Q = measure for the heterogeneity of the mean effect size; df = degrees of freedom in relation to the number of effects (k−1); p value = the probability of obtaining a value at least as extreme as the one that was actually observed, assuming that the null hypothesis is true.

our dataset of useable empirical (quantitative) studies was rather limited, the results of this review should be interpreted with caution. More explicitly, the heterogeneity of studies and the limited number of integrable effects constrain the practicability of our analysis. Still, we are quite able to derive some conclusions from our review.

Summing up all available data, it is safe to say that the results of this meta-analysis provided sufficient evidence that argues against the existence of the hot hand. The analysis of the effects yielded a minor positive overall mean effect size of .02. The present study provides additional support for Gilovich et al.'s 1985 claim that a general hot hand effect probably does not exist in sport. The possibility that the hot hand exists in specific type of sports has yet to be clearly demonstrated, since those studies claiming for the hot hand do not tend to extend beyond one or two published papers. Therefore, it is conceivable that these effects are not robust. The sport that yielded the most empirical results, basketball, does not reliably show a hot hand, which suggests that the case for other sports needs more evidences. Furthermore, given that our analysis was intentionally restricted to the sport domain, we obliviously wish to make no general statements about the hot hand and its origins in relation to other domains beyond sport.

Regarding the possibility of a publication bias, we believe that the best strategy for appraising evidence regarding the hot hand should be to consider the meta-analytic results alongside a (traditional) qualitative critical assessment of the studies in a systematic review. In any case, more than several additional studies with a positive effect or no effect at all (a zero correlation value) would be required in order to distort our results. In other words, only if a considerable number of studies that consistently show a similar effect are published would the general pattern of our findings be changed. It is quite unlikely to expect that such an incidence would occur, unless other research groups trying to replicate the very same study would report the same results.

As for our second goal, we were unable to extract any useful information that could explain the heterogeneity of effects in a satisfying way. We identified no particular environments in which the hot hand might possibly emerge. The cumulative evidence reflects scientific reality: the extremely heterogeneous distribution of effect sizes indicates that there is a huge random effects variation that should be explained before any statements regarding the existence of the hot hand can be made. Therefore, any further research should focus on aspects beyond the question of whether the effect under study exists or not.

Methodologically, there is only one way to support the existence of an effect – by trying hard enough to reject the assumption that it exists. Only an effect that can survive this procedure (i.e., can not be rejected), can be considered to be a real effect. Following this line of reasoning, most people should doubt whether the hot hand really exists, simply because the majority of studies have failed to support it. With modern statistics it is possible to detect a stable effect, even if each single study reports no significant effect. We found that in most cases the hot hand effect did not survive an empirical attempt to reject it. Even employing a meta-analytic tool, which is a fair strategy to detect even small effects, could not provide positive evidence.

Obviously, there is no final logical proof for the presence or absence of an empirical effect. Therefore, we will always find people who are not willing to consider the hot hand as merely an illusion. Amos Tversky used to describe the robustness of the belief in the hot hand by saying that he has been in thousand arguments over this topic, won them all, but convinced no one (Gilovich, 2002). Scientifically, the fact to be emphasized here is that we found no solid evidence for the existence of the effect across the domain of sports. Recent scientific developments relate the variability between the outcomes of repeated human performances to either limitations of our neuro-transmission system (Faisal, Selen, & Wolpert, 2008), inherent limit to human maintenance of concentration and attention (Stadler, 2007), or to factors such as shot selection or defense effort (Sun & Wang, 2012). In any case, people seem to act based upon improvisation by default (Churchland, Afshar, & Shenoy, 2006).

In light of these findings, we encourage researchers to seek a deeper understanding of the cognitive processes underlying pattern detection. Recent efforts to explore the mechanism of human judgment of random sequences were demonstrated by Hahn and Warren (2009) and Oskarsson et al. (2009). Wilke and Barrett (2009) attempted to explain the hot hand effect from an evolutionary point of view. They suggested that during human development the detection of streaks was advantageous for choosing appropriate living spaces. A streak should consist of at least three identical successive outcomes (three dimensions or three basic colors) in order to be noticed, due to either human learning processes or cognitive limitations (Carlson & Shu, 2007).

Further research should also address the consequences of the hot hand fallacy. For example, it would be instructive to investigate whether the illusion of being "hot" makes one a better player, even if it has no influence on the probability of succeeding in hitting the next shot. Another fruitful avenue for research would be to explore the allocation and replacement decisions made by players and coaches. This topic was partially explored by Gula and Köppen (2009) in volleyball. Hot hand sequences were manipulated by the parameters perfection and length. The results showed that players who had different levels of expertise were strongly influenced by the hot hand belief.

Concluding remarks

The contribution of the present study should be considered in light of its very exclusive attempt to address the challenge introduced by the existence of controversial answers to the question whether the hot hand in sports exists, and to discover under which circumstances it might survive. In spite of our efforts we found no solid evidence for either the existence of a general hot hand effect, or for any moderating variable that can explain the extent of a hot hand effect. Even though these challenges were not fully accomplished, this report should still be evaluated by its merit to provide guidelines for future research using recent new developments and potential issues that have been discussed here.

In light of the current available empirical and statistical evidence, we maintain that any further investment in a search for the hot hand is redundant unless the current state of the field is improved to include a larger number of highly methodological, quality studies. Other research groups are encouraged to derive specific quantitative predictions and to subsequently implement empirical studies that would allow conclusive assessment of the hot hand. In addition, hot hand research should be directed primarily to exploring the effects of the belief on decisions made by both athletes and coaches.

References

*Adams, R. M. (1992). The "hot hand" revisited: successful basketball shooting as a function of intershot interval. *Perceptual and Motor Skills, 74*, 934.

Adler, P., & Adler, P. A. (1978). The role of momentum in sport. *Urban Life, 7*, 153–176.

Albert, J. (2008). Streaky hitting in baseball. *Journal of Quantitative Analysis in Sports, 4*, Berkeley Electronic Press, Article 3.

Albert, J., & Bennett, J. (2001). *Curve ball: Baseball, statistics, and the role of chance in the game*. New York: Copernicus.

Albert, J., & Williamson, P. (2001). Using model/data simulations to detect streakiness. *The American Statistician, 55*, 41–50.

* Indicate studies included in the meta-analysis.

*Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association, 88*, 1175–1183.
*Arkes, J. (2010). Revisiting the hot hand theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports, 6*. http://dx.doi.org/10.2202/1559-0410.1198, Article 2.
*Avugos, S., Bar-Eli, M., Ritov, I., & Sher, E. The elusive reality of efficacy-performance cycles in basketball shooting: analysis of players' performance under invariant conditions, in preparation.
Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of "hot hand" research: review and critique. *Psychology of Sport and Exercise, 7*, 525–553.
Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (version 2) [computer software]*. Englewood, NJ: Biostat.
Burns, B. D. (2004). Heuristics as beliefs and as behaviors: the adaptiveness of the "hot hand". *Cognitive Psychology, 48*, 295–331.
Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand". *Psychonomic Bulletin and Review, 11*, 179–184.
Carlson, K. A., & Shu, S. B. (2007). The rule of three: how the third event signals the emergence of a streak. *Organizational Behavior and Human Decision Processes, 104*, 113–121.
Cheng, L. T. W., Pi, L. K., & Wort, D. (1999). Are there hot hands among mutual fund houses in Honk Kong? *Journal of Business Finance & Accounting, 26*, 103–135.
Churchland, M. M., Afshar, A., & Shenoy, K. V. (2006). A central source of movement variability. *Neuron, 52*, 1085–1096.
*Clark, R. D. (2003a). Streakiness among professional golfers: fact or fiction? *International Journal of Sport Psychology, 34*, 63–79.
*Clark, R. D. (2003b). An analysis of streaky performance on the LPGA tour. *Perceptual and Motor Skills, 97*, 365–370.
*Clark, R. D. (2005). Examination of hole-to-hole streakiness on the PGA tour. *Perceptual and Motor Skills, 100*, 806–814.
Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: empirical data from casinos. *Journal of Risk and Uncertainty, 30*, 195–209.
Dixon, M. J., & Pope, P. F. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting, 20*, 697–711.
Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician, 58*, 38–45.
*Dumangane, M., Rosati, N., & Volossovitch, A. (2009). Departure from independence and stationarity in a handball match. *Journal of Applied Statistics, 36*, 723–741.
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.
Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*, 292–303, doi:10.138/nrn2258.
*Frame, D., Hughson, E., & Leach, J. C. (2004). Runs, regimes, and rationality: the hot hand strikes back. Working paper.
Frohlich, C. (1994). Baseball: pitching no-hitters. *Chance, 7*, 24–30.
Gartzke, E., & Simon, M. W. (1999). Hot hand: a critical analysis of enduring rivalries. *Journal of Politics, 61*, 777–798.
*Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. *Psychonomic Bulletin and Review, 2*, 260–265.
Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: The Free Press.
Gilovich, T. (2002). *Transcript from Tom Gilovich hot hand homepage online chat*. http://www.hs.ttu.edu/hdfs3390/hh_gilovich.htm.
*Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.
Gould, S. J. (1989). The streak of streaks. *Chance, 2*, 10–16.
Gula, B., & Köppen, J. (2009). Einfluss von Länge und Perfektion einer Hot-Hand-Sequenz auf Zuspielentscheidungen im Volleyball [[Effect of length and perfection of a hot-hand sequence on allocation decisions in volleyball]]. *Zeitschrift für Sportpsychologie [Journal for Sport Psychology], 16*, 65–70.
Gula, B., & Raab, M. (2004). Hot hand belief and hot hand behavior: a comment on Koehler and Conley. *Journal of Sport and Exercise Psychology, 26*, 167–170.
Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: why three heads are better than four. *Psychological Review, 116*, 454–461.
Hales, S. (1999). An epistemologist looks at the hot hand in sports. *Journal of the Philosophy of Sport, 26*, 79–87.
Hancock, R., & Frisch, D. (2005). *The hot hand revisited*. Online document retrieved November 5, 2006 from. http://blog.1tc.arizona.edu/coblog/archives/2005/09/the_hot_hand_revisited_1.html.
Hastie, R. K., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making* (2nd ed.). California: Sage.
Hedges, L. V., & Olkin, I. (1985). *Statistical methods for metaanalysis*. San Diego, CA: Academic Press.
Hoffman, A. J. (1983). Effects of psychological momentum on the physiology and cognition among American athletes. *International Journal of Sport Psychology, 14*, 41–53.
Hooke, R. (1989). Basketball, baseball, and the null hypothesis. *Chance, 2*, 35–37.
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
Iso-Ahola, S. E., & Mobily, K. (1980). "Psychological momentum": a phenomenon and an empirical (unobtrusive) validation of its influence in a competitive sport tournament. *Psychological Reports, 46*, 391–401.
*Jordet, G., Hartman, E., & Vuijk, P. J. (2011). Team history and choking under pressure in major soccer penalty shootouts. *British Journal of Psychology, 103*, 268–283. http://dx.doi.org/10.1111/j.2044-8295.2011.02071.x.
Klaassen, F. J. G. M., & Magnus, J. R. (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., & Conley, C. A. (2003). The "hot hand" myth in professional basketball. *Journal of Sport and Exercise Psychology, 25*, 253–259.
Köppen, J., & Raab, M. (2012). The hot and cold hand in volleyball: individual expertise differences in a video-based playmaker decision test. *The Sport Psychologist, 26*, 167–185.
Korb, K. B., & Stillwell, M. (2003, July). The story of the hot hand: powerful myth or powerless critique? In *Paper presented at the international conference on cognitive science*. Sydney, Australia.
Lagnado, D. A., & Sloman, S. A. (2004). Inside and outside probability judgment. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 155–176). Oxford: Blackwell Publishing.
Larkey, P. D., Smith, R. A., & Kadane, J. B. (1989). It's okay to believe in the hot hand. *Chance, 2*, 22–30.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
*Livingston, J. A. (2012). The hot hand and the cold hand in professional golf. *Journal of Economic Behavior & Organization, 81*, 172–184.
*Miller, S., & Weinberg, R. (1991). Perceptions of psychological momentum and their relationship to performance. *The Sport Psychologist, 5*, 211–222.
Miyoshi, H. (2000). Is the "hot hands" phenomenon a misperception of random events? *Japanese Psychological Research, 42*, 128–133.
Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin, 135*, 262–285.
*Palacios, J. L. (2010). Runs of Markov chains and streaks in baseball. *Methodology and Computing in Applied Probability, 12*, 659–665. http://dx.doi.org/10.1007/s11009-009-9150-6.
Raab, M., Gula, B., & Gigerenzer, G. (2012). The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied, 18*, 81–94.
Reifman, A. (2012). *Hot hand: The statistics behind sports' greatest streaks*. Virginia: Potomac Books.
Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82.
Siwoff, S., Hirdt, S., & Hirdt, P. (1988). *The 1988 Elias baseball analyst*. New York: Collier.
Smith, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin and Review, 10*, 753–758.
Stadler, M. (2007). *The psychology of baseball: Inside the mental game of the Major League player*. New York: Gotham.
*Stern, H. S., & Morris, C. N.. (1993). A statistical analysis of hitting streaks in baseball: comment. *Journal of the American Statistical Association, 88*, 1189–1194.
Sun, Y., & Wang, H. (2012). The "hot hand" revisited: a nonstationarity argument. *PsyCh Journal, 1*, 28–39. http://dx.doi.org/10.1002/pchj.2.
Swartz, T. (1990). Letter to the editor: more on the "hot hand". *Chance, 3*, 6–7.
*Tversky, A., & Gilovich, T. (1989a). The cold facts about the "hot hand" in basketball. *Chance, 2*, 16–21.
*Tversky, A., & Gilovich, T. (1989b). The "hot hand": statistical reality or cognitive illusion? *Chance, 2*, 31–34.
Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases (Washington, D.C.). *Science, 185*, 1124–1131.
Tyszka, T., Zielonka, P., Dacey, R., & Sawicki, P. (2008). Perception of randomness and predicting uncertain events. *Thinking & Reasoning, 14*, 83–110. http://dx.doi.org/10.1080/13546780701677669.
*Vergin, R. C. (2000). Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior, 23*, 181–197.
*Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician, 49*, 24–28.
Wardrop, R. L. (1999). *Statistical tests for the hot-hand in basketball in a controlled setting*. http://www.stat.wisc.edu/~wardrop/papers/tr1007.pdf.
Wilke, A., & Barrett, H. C. (2009). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior, 30*, 161–169.
*Yaari, G., & Eisenmann, S. (2011). The hot (invisible?) hand: can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? *PLoS ONE, 6*(10), e24532. http://dx.doi.org/10.1371/journal.pone.0024532.