The Story of The Hot Hand: Powerful Myth or Powerless Critique?

Kevin B. Korb and Michael Stillwell
School of Computer Science and Software Engineering
Monash University, Clayton, Victoria 3168 Australia

(korb@csse.monash.edu.au)

Abstract

Thomas Gilovich and Amos Tversky famously claimed that the belief in the Hot Hand in basketball is a cognitive illusion, an instance of the belief in the Law of Small Numbers. They supported this claim with a variety of orthodox statistical tests failing to show any clear evidence of a Hot Hand phenomenon. These researchers neglected to perform any power analysis on this test. Here we show that their belief in having demonstrated the illusory status of the Hot Hand is itself an illustration of the Law of Small Numbers.

Introduction

An important and influential stream of research in cognitive psychology has been the study of human reasoning under uncertainty, and especially the study of cognitive illusions. Of particular importance for methodologists are Kahneman and Tversky's investigations of a fallacious belief in the "Law of Small Numbers" (Tversky and Kahneman, 1971), since a substantive implication is that much of the experimental work using orthodox statistical significance testing lacks statistical power. In an interesting thread of this research, Thomas Gilovich and Amos Tversky claimed that the belief in the "Hot Hand" in basketball is a cognitive illusion, an instance of the common, but erroneous, belief in the Law of Small Numbers (Gilovich *et al*, 1985; Tversky and Gilovich, 1989). They supported this claim with a variety of statistical tests failing to show any clear evidence of a Hot Hand. This idea that belief in the Hot Hand is illusory has since been widely promulgated, for example by Stephen Jay Gould in *Bully for Brontosaurus* (1991) and by Gilovich himself in his *How We Know What Isn't So* (1991). The curious fact is, however, that these psychologists themselves employed orthodox statistical significance tests, *without any power analysis.* Here we show that their belief in having demonstrated the illusory status of the Hot Hand is itself an illustration of the Law of Small Numbers, for their statistical tests were of such low power that they could not have been expected to find a Hot Hand even if it were present.

Background

Since Savage's seminal presentation of Bayesian statistics (Savage, 1954), the investigation of Bayesian inference has played a notable role in cognitive psychology. Ward Edwards was one of the first psychologists to investigate Bayesian reasoning as a normative standard against which the reasoning of humans could be compared (Edwards, 1954). Daniel Kahneman and Amos Tversky, in a series of articles (many collected in Kahneman, Slovic and Tversky, 1982), have demonstrated a large variety of ways in which ordinary forms of human reasoning diverge from Bayesian standards.

One of the more important contributions of this research program is Kahneman and Tversky's "Law of Small Numbers": the belief that many people have that small samples are generally just as representative of the population as are large samples (Tversky and Kahneman, 1971). This tendency reveals itself in many ways; one of the more interesting is the difficulty people have in recognizing or generating random sequences. Humans appear to be "hard-wired" to see patterns in noise. One Melburnian statistician of our acquaintance wins money from the foolhardy by challenging them to bet that random sequences which they create cannot be distinguished (with a blind control) from sequences produced by a random number generator: the statistician needs only to select those that *appear* improbably random, with fewer "streaks" of 1's and 0's and more alternations than are mathematically expected — or, to put it Kahneman and Tversky's way, to select those small samples that in superficial respects are improbably like large samples. No doubt, the urge (and ability) to see patterns where none exist has survival value in general and should not be glibly decried as "irrational". But it is also clear that in many cases hallucinating patterns and acting (e.g., betting) upon the hallucination is detrimental to economic (and other forms of) well-being.

Gilovich, Vallone and Tversky (1985; or, GVT) made a surprising application of the Law of Small Numbers: they claimed to demonstrate that the widely accepted belief that basketball players occasionally get into a "zone" where they are shooting better than normal is unfounded. This has been contentious for various reasons. Probably one reason is that, indeed, when commentators, coaches and fans see a player make six shots in a row, they tend to see a pattern, whether or not it is justified. Another reason is that intuitive introspection seems to reveal that such a thing as "the zone" in fact exists: it is nearly universal amongst those who do sports to report that on *some* occasions they have been in states where success can be anticipated, and not just noticed retrospectively. But introspection has come under fire within cognitive psychology as

at best unreliable and at worst as simply shoddy (Nisbett and Wilson, 1977). So, GVT collected a season's worth of basketball data, applied a variety of statistical tests that might uncover evidence of streak shooting, and, when these tests failed to reveal anything, they declared the non-existence of the "Hot Hand."

Whereas we agree with the underlying thesis that people have a tendency to see patterns when they do not exist and so are misled, and that this phenomenon is important, we believe Gilovich and company err in their sweeping dismissal of the Hot Hand. Although introspection is a more fallible research tool than many, in some cases introspective beliefs are hard sensibly to dismiss. And, in particular, we believe in a Symmetry Thesis: the Hot Hand exists if and only if the cold hand exists. Even Thomas Gilovich has conceded this much.[1] But the intuitive evidence for cold hands is overwhelming: it is as clear as anything that from time to time we play our sports worse than normal, if only because we become bad-tempered, or get a cramp, or what have you. But if occasional ordinary disadvantages can thus hamper play, it is exceedingly probable that occasional ordinary advantages will aid play. The plausible result is that, although over the long term there may be a fairly stable average ability with which anyone may play the game, this average is itself composed of some large variety of abilities, which are themselves sensitive to any number of factors that change over time (sleep, nutrition, emotional relationships, etc.). This corresponds to a statistical model of shots being drawn randomly from a series of distinct binomial distributions, each distribution having a different binomial parameter p governing the probability of making a shot, some high and some low, but with the stable average being the weighted average p value. This is our plausible alternative model to GVT's null hypothesis of a *single* binomial process; and, if it is true, then streak shooting should be observed occasionally, in both hot and cold varieties.

So, if streak shooting exists, and GVT examined a team's shooting record over an entire season (the Philadelphia 76ers over the 1980-81 season), then why didn't they find any streaks? GVT did take as their null hypothesis the simple binomial process (1985, p. 297): "The simple binomial model is equivalent to a more complicated process with the following characteristics: Each player has an ensemble of shots that vary in difficulty (depending, for example, on the distance from the basket and on defensive pressure), and each shot is randomly selected from this ensemble." This is correct. Gilovich *et al.* consider this significant since they believe the latter description is intuitively plausible for basketball shooting. Note that their statistical model is quite distinct from our alternative above, since our serial-binomial model does *not* suppose that each shot is randomly drawn from among all binomials in the ensemble; it is only from within a particular binomial distribution that a random selection is made on any given occasion. On either hypothesis, the null hypothesis of a single binomial process explaining shot data and the alternative of a series of binomial processes, the data available are simply sequences reporting hits or misses with basketball shots — on GVT's null hypothesis, indistinguishable from the outcomes of tossing a (possibly biased) coin.

A key point is this: binomial data are relatively impoverished. It is therefore quite difficult to extract meaningful information from them, requiring large samples. Despite the fact that GVT are working in a statistical tradition that takes its cue from Bayesian principles, they in fact chose to employ *orthodox* statistical tests — in particular, null hypothesis significance tests (NHST) of various kinds — in order to extract patterns from their data. NHST, as ordinarily conducted, frequently lack *statistical power* — a reasonable probability of reporting that an alternative hypothesis is true, when it is true — and that is a problem that is so well known, even among those trained as orthodox statisticians, that the American Psychological Association has called for power analyses to be standard in any experimental work published in its official journals (American Psychological Association, 1994). Kahneman and Tversky themselves, the intellectual progenitors of the Hot Hand study, denounced the neglect of power in null hypothesis significance testing, as a manifestation of a superstitious belief in the "Law of Small Numbers". Notwithstanding all of that, Gilovich *et al.* base their conclusion that the hot hand phenomenon is illusory squarely upon a battery of significance tests, having conducted no power analysis whatsoever! This is perhaps the ultimate illustration of the intellectual grip of the significance test over the practice of experimental psychology.

In contrast with GVT, we have performed power analyses of their significance tests. We have found that these significance tests are woefully inadequate to the task. Even if the Philadelphia 76ers did shoot in streaks, it is highly unlikely that Gilovich, Vallone and Tversky would have discovered that fact. In order to discover whether their data *in fact* hold patterns indicating some kind of streak shooting, we have also developed a *Bayesian* inference procedure for extracting such patterns from binomial data. Although it turns out that the basketball data do not offer any substantial support for the Hot Hand thesis, whether studied using orthodox tests or our new Bayesian tests, none of the tests have sufficient power for this to be an exciting result, or supporting to any significant extent the GVT proclamation that the Hot Hand phenomenon is illusory.

We conclude that the entire affair is best understood as demonstrating the need for improved statistical tools, and the need for greater caution in the employment of NHST, or better yet for the general abandonment of NHST for the purpose of testing substantial scientific hypotheses.[2]

The Gilovich-Vallone-Tversky Study

It is not entirely clear what the Hot Hand phenomenon is generally supposed to be, nor just what GVT intend us to un-

[1] In a personal communication. Perhaps it is worth noting that the Symmetry Thesis is clearly *not* an analytic truth — there are *logically possible* worlds where it is false. In asserting the thesis we are simply asserting our *a posteriori* conviction that we are not in such an extraordinary world.

[2] No doubt there have been many attempts to refute or deny the conclusions of Gilovich *et al* (1985). None that we are aware of has made our central point about statistical power. Wardrop (1995) is perhaps the most interesting, pointing out that believers in the Hot Hand may have been dealing only with data that aggregates basketball shooting across all players, rather than keeping them separate. This renders their interpretation subject to illusory support via Simpson's paradox. Whether this explanation is correct or not, however, it is directed at explaining human beliefs, rather than the correct interpretation of the GVT data.

Table 1: Serial correlation for Philadelphia 76ers, 1980-81; p value shown for statistically significant results.

	r_s	
Richardson:	-.020	
Erving:	.016	
Hollins:	-.004	
Cheeks:	-.038	
C Jones:	-.016	
Toney:	-.083	
B Jones:	-.049	
Mix:	-.015	
Dawkins:	-.142	$(p \leq .01)$

derstand by it.[3] Presumably, it is not meant to be so non-specific as to allow just *any* increase beyond the long-term average shooting ability, for they do not seem to be denying the effects of training, maturation, etc. In contrast with such long-term effects, in common parlance it seems that the presence of a Hot Hand is meant to be unstable, disappearing within perhaps minutes but in any case within some weeks.

Gilovich *et al.* employed three significance tests in their search for the hot-hand:[4]

1. *Serial Correlation (r_s).* This measures the tendency of one shot to influence the next shot. If $r_s > 0$, then scoring on one occasion will tend to lead to scoring on the next. On the basketball data kindly supplied to us by Gilovich, we generated Table 1, which is substantially the same as GVT's Table 2 (p. 301).[5] The conclusion is that only one player shows a statistically significant deviation in the serial correlation among his shots, and that correlation is *negative*, suggesting some tendency for hits to be followed by misses and *vice versa*. Either a hot hand or a cold hand, if revealed in serial correlation, would lead to a positive correlation: a hot hand would tend to produce hits being followed by a higher probability of more hits; a cold hand would tend to produce misses being followed by a higher probability of further misses. In the serial correlation test we find no support for the Hot Hand hypothesis and, possibly, some evidence against it in the performance of Dawkins.

2. *Runs test.* The more streaky one's shooting is, the fewer alternations there will be between sequences of 1's or 0's, so the fewer "runs" of 1's and 0's will be found.[6] Given a single binomial process, one can compute an expected number of runs for the sample size. Table 2 shows the Z statistic for the actual number of runs. The only statistically significant results tend to support the claim that there are too many runs for a single binomial process, whereas the Hot Hand thesis would lead us to expect the opposite.

Table 2: Z statistic (number of standard deviations from null) for the runs test for Philadelphia 76ers, 1980-81; p values shown for statistically significant results.

	Z	
Richardson:	1.2954531	
Erving:	-0.53428394	
Hollins:	-0.6249232	
Cheeks:	0.8464658	
C Jones:	-0.22160907	
Johnson:	0.54589355	
Toney:	1.967824	$(p \leq .05)$
B Jones:	1.2091007	
Cureton:	-0.55472046	
Mix:	-0.09273166	
Dawkins:	3.0895486	$(p \leq .01)$

Table 3: χ^2 test p values for the Philadelphia 76ers, 1980-81.

	p
Richardson:	0.536
Erving:	0.940
Hollins:	0.873
Cheeks:	0.671
C Jones:	0.790
Johnson:	0.895
Toney:	0.070
B Jones:	0.477
Cureton:	0.995
Mix:	0.643
Dawkins:	0.122

3. *Chi-square test.* GVT also divided their shot data into successive groups of four. If these data were produced by a single binomial process (with the binomial parameter near 0.5, as indicated by the average hit rate), then the expected number of such groups with zero, one, two, three or four can be easily computed, with two or three hits predominating. On the other hand, if hot or cold hands occur with any frequency, then the number of groups with zero or one hit, or again with three or four hits, will be greater. GVT performed a χ^2 test to determine whether the observed frequency counts deviated significantly from what would be expected from a single binomial process. The p values are reported in Table 3, and none of them are statistically significant. Once again we fail to find any evidence of a Hot Hand in operation.

An obvious rebuttal to this sort of study is the suggestion that, although Hot Hands really do exist, opposing teams can recognize them as readily as anyone, and that they respond by closer guarding, or double guarding, so that the hot player's raised ability to hit the basket is negated. The result could easily be that the hot player's strike rate is unchanged. GVT

[3] Wardrop (1999) makes a strong case that GVT are in fact inconsistent in their characterization of the hot-hand thesis, sometimes describing it as a tendency for a hit (miss) to raise (lower) the probability of a subsequent hit and sometimes as a short-term increase in the binomial parameter governing the probability of a hit (non-stationarity). We tend toward the latter interpretation as a more plausible rendition of what athletes and others have reported.

[4] They employed one additional test, which looked for variation in play from game to game. Although we were kindly provided with a photocopy of the original data, this did not include an indication of when games began or ended, so we have ignored this test.

[5] Minor deviations from their reported statistics may be due to minor corruption of the data during photocopying. The same is true of the other statistical tests performed on the real data, below.

[6] In other words, a "run" is an uninterrupted sequence of 1's or an uninterrupted sequence of 0's. Given either hot or cold hands, there will be fewer of these.

have pre-empted such a move by also conducting a controlled test, measuring free throw shooting results, when opposing team behavior is irrelevant. Statistical tests on these data again showed no statistically significant difference from the null hypothesis of a single binomial process.[7]

In summary, GVT have performed three common statistical tests on real basketball data; each test might be expected to reveal that these shot data were *not* generated by a single binomial process, and so, to be explained, require some more complex hypothesis, and in particular the invocation of some variety of Hot Hand. But none of them did. GVT conclude that the belief in a hot hand phenomenon in basketball, despite being nearly universal amongst players, coaches and fans, is illusory — "The Hot Hand as Cognitive Illusion" is, in fact, nothing more than a manifestation of an underlying belief in the Law Small Numbers (Tversky and Gilovich, 1989).

Power Analysis

It is highly meritorious of Gilovich, Vallone and Tversky to have gone to the trouble to collect and examine real basketball data in order to test the Hot Hand thesis. On the other hand, it is at best a curious choice for them to select *orthodox* statistical tests for inferring whether or not a Hot Hand phenomenon is real. These are psychologists working explicitly within a Bayesian tradition in psychology: Daniel Kahneman and Amos Tversky in particular made themselves famous by examining human inference under uncertainty and comparing such performance with a Bayesian standard. Bayesian statistics is well known to be antagonistic to orthodox statistical procedure in general and significance testing in particular. Statistical significance tests fail to take into account the prior probabilities of hypotheses under test. Even more tellingly, they fail to take into account the *power* of a test: the probability that the test will reveal a positive effect when there is one. In consequence, many experimenters radically overestimate the probability that a significance test, once it has come out positive, would do the same on a repetition. This is once again a manifestation of a belief in the Law of Small Numbers: in their seminal paper "Belief in the Law of Small Numbers" Tversky and Kahneman (1971) castigate their peers for employing significance tests in this fashion, that is, while neglecting the power of their tests.

The mathematical analysis of power can be difficult or impossible. In particular, when there is a continuum of hypotheses alternative to the null hypothesis, any given test will have associated with it a continuum of power values. Commonly, researchers select a "least interesting" alternative hypothesis, one whose effect size is as small as possible while nevertheless being substantial enough to make a credible practical or theoretical alternative to the null effect, and compute the power of the test relative to that alternative. We have done that for GVT's significance tests. We ran GVT's tests on three cases of artificially generated data, in each case for a total of 110 times (ten times for each of eleven basketball players). Each of the 110 binomial sequences corresponded to an actual sequence in length and frequency of hits. Two of the cases present data generated by the serial binomial alternative to GVT's null hypothesis; that is, they were generated by models of the Hot Hand hypothesis. Thus, if GVT's tests have high power relative to these alternative models, they should be able to discover hot hands a reasonably large percentage of the time. The three cases produced were:

1. Null basketball data. These were produced by a single binomial process with p set to the player's actual hitting frequency in the data (i.e., these were independently identically distributed sequences, without any split). Any hot hand discovered in this case is a type I error.
2. One-split basketball data. Two binomial processes produced the data: the first half governed by the binomial parameter $p-0.05$ and the second half by $p+0.05$, where p is the actual frequency of hits in the data. For example, for a player with an overall hitting frequency of 0.5 the first half of the shots were generated with $p = 0.45$ and the second half with $p = 0.55$.[8]
3. Four-split basketball data. Two binomial processes generated the data, one producing two short (10 shot) streaks around the 1/3 and 2/3 points of the sequences and with $p = 0.9$, with the actual frequency of hits used for the remainder. An example graph of the p parameters for a generating sequence of binomial distributions is shown in Figure 1.

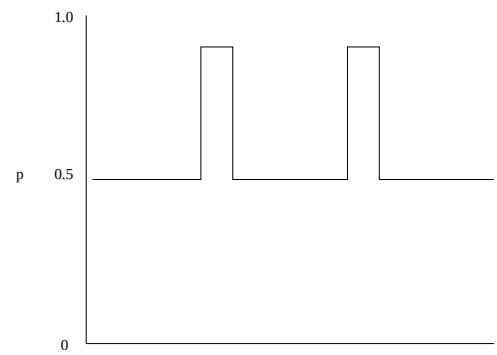

Figure 1: Binomial parameters for four-split data

Table 4: Frequency of one or more splits found on artificial data

Splits	r_s	runs	χ^2
0	.04	.04	.15
1	.07	.08	.19
4	.07	.08	.22

Table 4 reports the frequency with which GVT's tests reported a significant result favoring the Hot Hand hypothesis — that is to say, favoring the presence of one or more splits in the data. The first row shows a meritorious refusal to report the existence of a Hot Hand when none exists, although χ^2 is a little loose on this point. The remaining two rows show that in neither alternative scenario, when some variety of Hot Hand indeed is behind the data, can GVT's tests discover the

[7] These data were not supplied to us. Wardrop (1999) examined the statistical tests employed on these data by GVT and concluded that the tests were also of low power.

[8] As Wardrop (1999) points out, a swing of 0.1 in shooting ability is very substantial for a professional basketball player. In any case, we conducted the same test with p values up to 0.40, with results in line with what we report here.

fact with any reliability. At best, χ^2 correctly identifies such cases around 20% of the time. In other words, even if the Hot Hand phenomenon were real, we should hardly expect Gilovich *et al* (1985) to have found evidence of it. They have wasted their time and ours. In promoting the view that the Hot Hand is a cognitive illusion, they have succumbed to their own meta-illusion, confusing the report of a weak statistical test applied to a small sample as representative of an enduring and significant phenomenon.

Bayesian Inference of Streaks

Beyond pointing out this substantial flaw in the interpretation of their experimental work, we would ideally come up with an alternative statistical test with high power, apply the test to the data, and discover that, contrary to all reports, there is indeed evidence of a Hot Hand to be extracted from the raw data after all. We have, in fact, taken the first steps, developing a Bayesian test for the Hot Hand, and indeed one with somewhat more power than GVT's orthodox tests.

Our Bayesian test uses an information-theoretic criterion designed specifically to test against each other models of sequences of binomial processes. In principle it does this by generating and comparing posterior probabilities for the null hypothesis of a simple binomial process, for all sequences of two binomial processes (differentiated by the location in the data where the second binomial process takes over, known as the "cutpoint" in the data), for all sequences of three binomial processes, etc. In practice, the Bayesian test stops when adding an additional cutpoint fails to deliver any improvement in posterior probability over models with one fewer cutpoint. The Bayesian test inherently favors simplicity over complexity; thus, if the null hypothesis does as well as any alternative in fitting the data, the null hypothesis will be preferred. Note that our test has a substantial advantage over the orthodox tests supported by Gilovich *et al.*: if and when the Bayesian test supports an alternative to the null hypothesis, it will report back the number and locations of the cutpoints found. Orthodox significance tests can only report that the null hypothesis has been rejected, leaving open what alternative hypothesis might be true.

Our Bayesian test for sequencing binomial data was developed using the information-theoretic inferential method of Minimum Message Length (MML). MML inference is a highly successful Bayesian method of computational data analysis, developed initially by Wallace and Boulton (1968).[9] The fundamental idea behind MML is to assess the probabilities of hypotheses by the lengths of messages required to communicate them, following Shannon's account of information, which asserts: $\text{Message Length}(h) = -\log P(h)$ — i.e., the optimal message length, minimizing the entropy over a communication line, is the negative log of the probability of the message being sent. Coding theory provides methods to encode known probability distributions so as to satisfy Shannon's equation. Given that, a search of the hypothesis space for the minimum length message is equivalent to a search for the hypothesis with the highest posterior probability, since $\min(\text{Message Length}(h)) = \max(\log P(h))$. MML has been successfully applied to problems as diverse as unsupervised clustering and mixture modeling (Wallace and Boulton, 1968), linear regression (Baxter and Dowe, 1994), supervised classification (Wallace and Patrick, 1993), and the discovery of causal models (Wallace and Korb, 1999).

Our application of MML to binomial sequences is original and has led to subsequent work on such segmented processes (Viswanathan *et al.*, 1999). The details of the MML code for binomial sequences we developed for this study are in (Stillwell, 1998). Our MML algorithm, per above, searches the space of hypotheses consisting of sequences of binomial distributions, from simpler to more complex. It returns the binomial sequence, including cutpoints, which is most probable, given the data available. Hence, given the actual basketball data collected by Gilovich *et al* (1985), or the data artificially generated by us, MML discovers the hypothesis concerning binomial sequences most likely to have generated that data. Of course, like any consistent (convergent) statistical inference procedure, it will perform better given larger sample sizes; identifying the sample sizes needed just is power analysis.

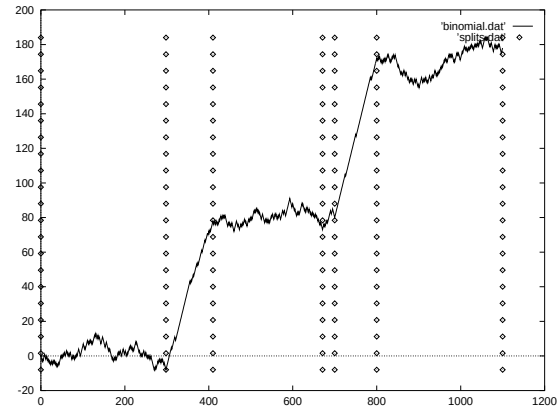

Figure 2: MML segmentation example

As an example, consider Figure 2. This reports a binomial sequence in cumulative fashion, as a kind of random walk with hits being represented by 1 and misses by -1. The sequence was actually generated by five binomial processes, with short segments at the 1/3 and 2/3 points using parameter $p = 0.9$ and the others with $p = 0.5$. MML discovery had no trouble finding these sequences, although it did also find a spurious cutpoint immediately prior to the second "Hot Hand".

Table 5: Frequency of one or more splits found on artificial data

Splits	r_s	runs	χ^2	MML
0	.04	.04	.15	.16
1	.07	.08	.19	.49
4	.07	.08	.22	.15

Having done a power analysis of the weak tests used by Gilovich *et al* (1985), we of course also performed a power analysis of our own test. The results are reported in Table 5, which extends the prior analysis to incorporate MML on the same tests. These results are not actually very encourag-

[9] It should not be confused with Minimum Description Length inference (MDL), which was developed by Rissanen (1978) in response to Wallace and Boulton (1968) and specifically as an *anti-Bayesian* inference method.

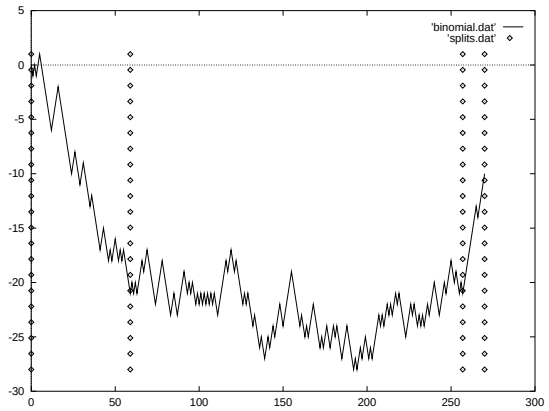

Figure 3: Caldwell Jones data

ing: on the data set sizes actually recorded by Gilovich *et al* (1985), MML never even reaches a 50% chance of finding a binomial split when there is one. When we applied MML discovery to the actual data, as opposed to the artificial data for power analysis, we obtained a consistent result. That is, MML found two of 16 players who took more than 100 shots during the year had split points. One such was Caldwell Jones, as shown in Figure 3. The Jones case looks quite promising, since although the last split looks a good deal like an accident at the end of the year, the first looks like a real recovery from a poor start — and so, an example of recovery from a cold hand. But before getting very excited by this kind of case, we need to consider the probability of MML finding false positives, reporting splits that are unreal. The frequency of false positives in the simulated data (the first row in Table 5) is 0.16. Given 16 basketball players, we have an expected number of false positives equal to 2.56; and MML reported three splits. These results are clearly consistent with there being no real splits in the basketball data.

Conclusion

Does this mean that GVT were right after all? Is the Hot Hand in basketball a mere illusion? That sort of conclusion is justifiable only under the assumption that the probability of the various tests, MML included, of recovering splits that are real — the power of the tests — is reasonably high. But we have seen that all of the tests have quite low power, in no case reaching even 50%. The reality is that binomial data, in contrast with most varieties of data, are statistically impoverished: it takes much larger samples to obtain useful estimates of the underlying process than with richer data, such as sample points from a normal distribution. Gilovich *et al* (1985) were incongruously conservative in employing orthodox significance tests to investigate the Hot Hand phenomenon, rather than search for suitable Bayesian inferential procedures. They were surprisingly neglectful in failing to perform any kind of power analysis for the statistical tests they did select. Doing so would have revealed just how weak their tests were, and so also the need for much larger sample sizes than were collected. But worst of all was their claim to have demonstrated the Hot Hand to be a cognitive illusion. In this they succumbed to the Law of Small Numbers themselves, for they were supposing that small, information-poor samples of basketball shooting should support the very same inferences that an arbitrarily large sample could. Nearly a generation of researchers and commentators has subsequently taken it for granted that the Hot Hand phenomenon is non-existent, and not just in basketball but quite generally. A serious reason to believe that negative conclusion has yet to be offered.

Acknowledgements

We wish to thank members of the Melbourne Reformasi and Susan Rogerson.

References

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington: APA.

Baxter, R.A. and Dowe, D.L. (1994). Model selection in linear regression using the MML criterion. *Proceedings of the Fourth IEEE Data Compression Conference*, Snowbird, Utah, p. 498.

Edwards, W. (1954). The reliability of probability preferences. *American Journal of Psychology, 67*, 68-95.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* NY: The Free Press.

Gilovich, T., Vallone, R. and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295-314.

Gould, S.J. (1991). *Bully for Brontosaurus: Reflections in Natural History.* NY: W.W. Norton.

Kahneman, D., Slovic, P. and Tversky, A. (eds.) (1982). *Judgment under uncertainty*. Cambridge University.

Nisbett, R.E., and Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*, 465-471.

Savage, L.J. (1954). *The foundations of statistics*. NY: Wiley.

Stillwell, M. (1998). *Investigating the "Hot Hand" Phenomenon.* Honours Thesis, Computer Science, Monash University.

Tversky, A. and Gilovich, T. (1989). The cold facts about the "hot hand" in basketball. *Chance: New Directions for Statistics and Computing, 2*, 16-21.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 2*, 105-110. Reprinted in Kahneman *et al.*, 1982.

Viswanathan, M., Wallace, C.S., Dowe, D.L. and Korb, K.B. (1999). Finding cutpoints in noisy binary sequences: A revised empirical evaluation. In N. Foo (Ed.) *Lecture notes in artificial intelligence, 1747*, 405-416.

Wallace, C.S., and Boulton, D.M. (1968). An information measure for classification. *The Computer Journal, 11*, 185-194.

Wallace, C.S. and Korb, K.B. (1999). Learning linear causal models by MML sampling. In A. Gammerman (ed.) *Causal models and intelligent data management.* Springer Verlag.

Wallace, C.S. and Patrick, J.D. (1993). Coding decision trees. *Machine Learning, 11*: 7-22.

Wardrop, R.L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician, 49*, 24.

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`