

Streaks in skilled performance

DAVID L. GILDEN

University of Texas at Austin, Austin, Texas

and

STEPHANIE GRAY WILSON

Seton Hall University, South Orange, New Jersey

Experiments in golf putting and darts demonstrated that skilled performance is streaky. The tendency for outcome sequences to form streaks was greatest when the task difficulty was such that about half the trials were successful. Mixtures of the two activities were also streaky, even when periodic interruption made the individual components resemble a random Bernoulli process. Formal models of sequence structure revealed that waves in hit rate are associated with the appearance of streaks.

There is a common perception shared by athletes, sports fans, and perhaps anybody who has ever attempted a skilled activity that there are moments when a person is "hot," or "in the zone," or alternatively is "cold," or in a slump. M. Csikszentmihalyi (1975, 1990; M. Csikszentmihalyi & I. S. Csikszentmihalyi, 1988) has written extensively about the notion of *flow* as a dimension of experience where the intimate coupling of actor and activity results in a specific psychological state anecdotally referred to as being "hot." Other investigators have been leery of flow as a description of performance and have questioned its empirical support. An issue germane to studies of subjective probability is whether the alleged signature of the flow state, streaky performance, arises from some factual aspect of skilled activity or from mistaking chance fluctuations as evidence of a heightened state of ability. Judgments of one's own and other's performance would not be controversial if it were not for the well-documented observation (Tversky & Kahneman, 1971, 1974) that people have systematic biases concerning what a random process looks like, and therefore are not competent to discriminate chance occurrence from truly remarkable performance.

Understandings of random processes incorporate a bias known as the "law of small numbers" (Tversky & Kahneman, 1971); people expect statistical regularities that hold only in the long run to also obtain locally in the short run. The gambler's fallacy (e.g., the belief that tails are "due" after a long run of heads) illustrates this type of thinking. When asked to mimic the flip of a fair coin, people invariably generate sequences that have too few long runs of heads and tails, they balance the frequencies of heads and tail in short runs, and they produce an excessive number of runs compared with the output from a

Bernoulli process (Kubovy & Gilden, 1989). Given this strong cognitive bias, it would not be surprising if people mistook a real Bernoulli process for streaky performance.

The game of basketball provides a singular opportunity to study the perception of skilled performance, since fans, coaches, athletes, and announcers all consider streaks to be a factual part of the game. Gilovich, Vallone, and Tversky (1985) compared individual shot records of professional NBA players with the expectation of a stationary Bernoulli process—a process characterized by the invariance of the probability of a successful field goal. Evidence for streak shooting (hot or cold) would be a positive correlation between outcomes of successive shots. Gilovich et al. (1985) found that individual shooting records did not support the existence of a "hot hand." In fact, the data supported the opposite conclusion; outcomes were negatively correlated (hits are more likely to follow misses than hits). This analysis implies that the widespread belief in streak shooting is due to misperception of chance.

Larkey, Smith, and Kadane (1989) criticized this conclusion on the grounds that a Bernoulli process does not capture important contextual features of event outcome in basketball. Defensive interference, pauses in the action, and variations in the opportunity to manifest the required skills together contribute to nonstationarity in the probability of a successful shooting attempt. Larkey et al. argued that observers of the game are highly sensitive to context and do not base their judgments of performance on individual shot records abstracted from the potpourri of activity that constitutes the spectacle.

Lacking in this discussion has been a consistent experimental paradigm that permits a rigorous demonstration of the phenomena. In this regard, basketball is manifestly not the optimal activity for analysis. Indeed, any game that is sufficiently interesting to draw spectators may be too complex for a rigorous streak analysis. We have chosen to study games that do not incorporate the stochastic variables of shooting opportunity and interfer-

We wish to extend special thanks to Kerstin MacDonald for assistance in running the experiments and preparing the data for analysis. Correspondence should be directed to D. L. Gilden, Department of Psychology, Mezes 330, University of Texas, Austin, TX 78712 (e-mail: gilden@psyvax.psy.utexas.edu).

ence, and so are able to enormously simplify the problem of selecting an appropriate normative model of chance. In our studies, people simply repeat a set task for a predetermined number of trials. Although rote repetition substantially vitiates the factors that impel people to exercise their abilities, the Bernoulli process generates the natural sampling distribution under these conditions.

The primary purpose of this article is to develop an experimental framework for investigating the existence of the "hot hand." We address the relatively straightforward empirical question of whether the serial execution of a skilled action is distinguishable from a Bernoulli process when the playing conditions are identical on all trials. We selected two motor skills that seemed to offer the opportunity for streaky performance: golf-ball putting and dart throwing. These games have the virtues that there is minimal delay in the action, there is no interference, there is an unequivocal criterion of success, and there is anecdotal evidence that these activities produce streaks.

INITIAL GOLF PUTTING STUDY

Method

Subjects. Forty subjects were recruited by advertisement. They received \$5 per session plus 5 cents per hit.

Stimuli. A standard 12-ft putting green, putter, and regulation golf balls were used.

Design and Procedure. Each subject completed one session of 300 putts following a brief practice session of 25 putts. Putting was initiated for all subjects at the far end of the putting green. The experimenter was placed near the hole and continuously resupplied the subject with fresh balls. Trials were self-paced. Hits and misses were encoded as 1s and 0s, respectively.

Analysis. There are a number of sequence statistics that measure deviation from the output of a Bernoulli process. Gilovich et al. (1985) used conditional probabilities, run counts, and serial correlations to characterize the basketball sequences in their studies. These measures are not independent. For a given hit rate, sequences with fewer runs than expected under the null hypothesis of Bernoulli trials must have more internal repetition than expected, and consequently a positive serial correlation. In addition, for such sequences, $p(h|h) > p(h)$ —the probability of a hit following a hit is greater than the probability of a hit.

Unlike the serial correlation between successive trials, denoted here as r_{12} , and the contingent probability difference, $\Delta p = p(h|h) - p(h)$, the hit rate influences the expected number of runs. A sequence with a hit rate near 1.0, for example, must have very few runs, although this does not mean that the sequence is streaky. Removal of hit rate as a factor in a runs analysis is accomplished by referring the run count in an observed sequence to the ensemble of counts calculated from all possible permutations. The sampling distribution of run counts is approximately normal with mean $2Np(1-p)+1$, where N is the number of trials and p is the probability of a hit. Deviations from normality are sufficiently large for $p \neq .5$ for us to use the exact hypergeometric distribution (Hays, 1988) to compute the probability of observing R , or fewer, runs. For the purposes of statistical testing and displaying our results, we have converted this probability to a z score by inverting the cumulative Gaussian distribution—a quantity that we refer to as the *runs z score*. The runs z score is a measure of outcome clustering that is independent of both sequence length and hit rate, and its expectation for a Bernoulli process is zero.

Although r_{12} , Δp , and the runs z score are related measures of sequence structure, they are not identical. The Fisher Z associated with r_{12} and the runs z score are numerically indistinguishable to three significant digits. Δp , however, is not a function of the runs z score. Simple regressions of these two variables captures only about 50% of the variance for the sequences discussed in this article. The reason for this is

that the first-order transition probability $p(h|h)$ is not as sensitive to global structure as is the serial correlation or run count. It is necessary to look at the first-, second-, and third-order transition probabilities to begin to adequately characterize a sequence. Often sequences with large disparities in their runs z scores differ only subtly in their transition probabilities. Transition probabilities are more useful as measures of local sequence structure, and for this reason we shall use the runs z score to characterize the departure from a Bernoulli process.

Runs z scores are subjected to two types of statistical tests, depending on the type of null hypothesis that is being considered. In every experiment reported here, we shall test whether the ensemble of outcome sequences is distinguishable from a Bernoulli process. The null hypothesis that sequences are samples from a Bernoulli process entails (1) each sequence's being independent (as are parts of sequences) from all others, and (2) the ensemble of runs z scores having a mean of zero. In testing this null, we simply form the distribution of observed runs z scores and ascertain whether the mean is significantly less than zero (we are only interested in the case where there are fewer runs than expected). Each sequence in such an analysis forms a separate degree of freedom by virtue of the mutual independence required by the null. We also test whether the runs z scores at different hit rates have different means. Here the null hypothesis has nothing to do with whether or not the sequences are Bernoulli, and we shall resort to more traditional repeated measure analyses where the different subjects comprise the degrees of freedom.

Results

The results for this experiment are shown in Figure 1. The distribution of runs z scores had a mean of $-.49$, which is distinguishable from a Gaussian with a mean of zero, the expected distribution derived from the null hypothesis of sampling from a Bernoulli process [$t(39) = -2.69$, $p < .005$]. Most people showed some amount of streakiness in golf putting (25 of 40 had $z < 0$; $p = .077$). In addition, there were a number of people who had runs z scores that were sufficiently negative for them to be regarded as streak performers. Of the sequences, 12.5% had runs z scores less than -2 , where only 2.3% are expected to have z scores this negative. These data also suggest that there is a suppression of streaky performance if the task difficulty exceeds the person's competence—that is, if the hit rate was .3 or less.

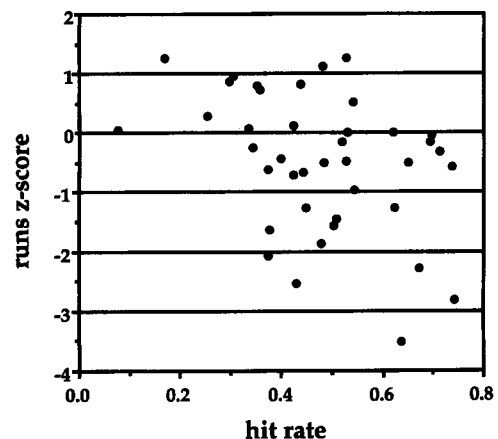


Figure 1. z scores for run number versus hit rate for 40 subjects in a golf putting study. Runs z scores are computed for each sequence individually on the assumption that successive trials are samples from a Bernoulli process defined by the average hit rate. Negative runs z scores indicate that fewer runs were observed than expected.

GOLF PUTTING STUDY THREE-LEVEL HIT RATE

A second putting experiment assessed the implied dependence of streakiness on hit rate by systematically varying the putting distance for each subject.

Method

Subjects. Five subjects were recruited by advertisement. They received \$5 per session plus 5 cents per hit.

Stimuli. The 12-ft putting green, putter, and balls were the same as those used in Experiment 1. The only difference in this experiment was that the subjects often initiated their putts some additional distance beyond the end of the putting green on a carpeted floor.

Design and Procedure. Each subject completed three sessions of 300 putts at each of the three levels of difficulty. Prior to each putting session, we calibrated the putting distance for each subject to achieve hit rates in the target ranges of .3-.5, .5-.7, and .7-.9. The difficulty level was chosen at random for each block of trials with the constraint that each subject complete three blocks at each level. Otherwise, the procedure was identical to that of the first putting experiment.

Results

The ensemble of sequences generated in this experiment had a mean runs z score of $-.59$. This is significantly smaller than zero [$t(44) = -3.94, p < .0001$], implying a runs deficit relative to the expectation of a Bernoulli process. There was additional structuring by hit rate observed in the runs z scores. Box plots of runs z score by difficulty level are shown in Figure 2. As anticipated, we found a suppression of streakiness in the low-hit-rate condition; the distribution of runs z scores in the range [.3-.5] had a median closer to zero, the expected value for independent trials. There also appeared to be a complementary and weaker suppression when the task was relatively easy. A repeated measures analysis showed that the U-shaped trend evident in Figure 2 was significant [$F(1,8) = 9.558, p < .02$].

The existence of a quadratic trend in runs z score with hit rate is a potentially important result. Outcome sequences are streakiest when the difficulty of the task is

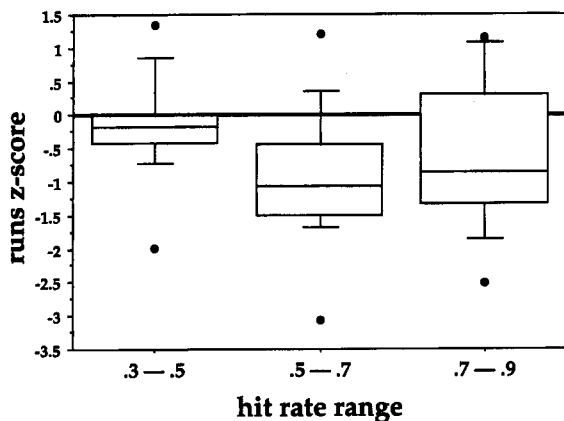


Figure 2. Box plots of runs z score versus hit rate in a golf putting study. The data are grouped according to the level of difficulty. Horizontal lines in the box plots mark the 10th, 25th, 50th, 75th, and 90th percentiles. Outliers are plotted as individual points.

commensurate with the ability of the performer. In the instruction of a skill, there is the commonsensible idea that learning is facilitated when difficulty is matched to ability, and it is surprising to see this notion instantiated in a formal measure of streak magnitude. We have assessed the generality of this result by repeating the three levels of hit-rate design in dart throwing. Darts is one "sport" where there is a folklore of streak shooting, although the documentation for this is presently limited to eyewitness accounts.

DART THROWING STUDY: THREE-LEVEL HIT RATE

Method

Subjects. Eight subjects were recruited by advertisement. They received \$5 per session plus 5 cents per hit. All of them considered themselves to be well-practiced and highly trained.

Stimuli. A standard Nodor dart board was placed at the regulation height of 5 ft 8 in. Subjects threw darts from a distance of 8 ft measured from the back of the board. The target for the calibration phase was a series of concentric alternating black and white rings that were $\frac{1}{2}$ -in. thick. During the test phase, the target was a filled black circle on a uniform white field. These targets were taped onto the Nodor dart board and were replaced after each set of 25 darts to facilitate accurate counting of hits.

Design and Procedure. We determined the dart throwing ability of each subject during an initial phase in which they threw 300 darts at a special calibration target. These data served to define individual calibration functions tabulating the number of hits within a radius, r , on a grid spaced at $\frac{1}{2}$ -in. intervals. The appropriate sizes of targets for achieving hit rates in the ranges, [.3-.5], [.5-.7], and [.7-.9], were interpolated from splines of the calibration functions. In the subsequent experimental sessions, hit-rate levels were chosen at random and subjects were given targets consistent with their calibrated abilities. The subjects completed one session of 300 throws at each of the three levels of difficulty. Trials were self-paced, as in the golf study, but safety dictated that the subjects retrieve their own darts. The subjects used either 3 or 5 darts depending on whether they brought their own or used those supplied.

Results

The data from this experiment are summarized as box plots of runs z score in Figure 3. A repeated measures analysis showed that the quadratic trend in runs z score with hit-rate range was significant [$F(1,14) = 7.26, p < .02$]. Dart throwing, however, did not generate solid evidence of streakiness per se. Of the 24 sequences generated in this experiment, only 1 had a runs z score sufficiently negative ($z = -1.6, p < .05$) to be regarded as anomalous. Chance mixing of outcome is expected to generate about 1 z score in 20 this negative. The mean runs z score was 0.11, not significantly different from zero [$t(23) = 0.47, p = .68$].

Discussion

A plausible account of the U-shaped functions must take into account what a person is doing when the task is too easy or too hard. At low hit rates, the performer manifestly does not possess the control to execute the required task. Successful trials must become increasingly random events as the hit rate approaches zero. At the other extreme, when the task is too easy, intermittent distraction may

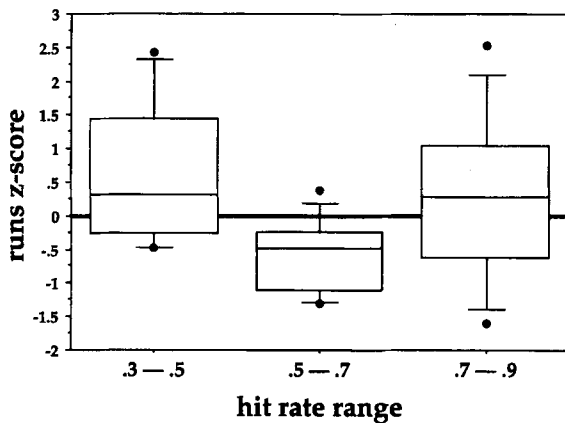


Figure 3. Box plots of runs z score versus hit rate in a dart-throwing study. The data is grouped according to the level of difficulty.

produce random failure. When successes or failures are truly random, then the runs z scores should approach the expectation value of zero. In this way, positive sequential dependency is most likely to manifest itself at intermediate hit rates by virtue of the true randomness that exists near ceiling and floor. Analysis of the dart data supports this argument. Runs z scores for sequences derived from the easiest (largest) and most difficult (smallest) targets were not distinguishable from a Bernoulli process. For the largest targets, the mean runs z score was 0.29 [$t(7) = .63, p = .55$]. For the smallest targets, the mean was 0.62 [$t(7) = 1.58, p = .16$]. Streaky performance was evident only for intermediate-size targets, where the mean runs z score of -0.58 was significantly different from zero [$t(7) = -2.88, p < .02$]. The golf data are in partial agreement with this interpretation. At the longest putting distances, where hit rates were lowest (hit rate $< .5$), the mean runs z score of -0.11 was not significantly different from zero [$t(14) = -1.02, p = .32$]. However, unlike the dart data, the high-hit-rate (hit rate $> .7$) golf sequences had a mean runs z score of -0.66 that is inconsistent with output from a Bernoulli process [$t(14) = -2.97, p < .01$]. Still, at putting distances where hit rates were in the range $[.5-.7]$, the mean runs z scores was -0.99 , the most negative value obtained in this set of studies [$t(14) = -6.4, p < .0001$]. We tentatively interpret the U-shaped functions as demonstrating that intermittency may make outcomes more Bernoulli-like at the endpoints of the hit-rate range.

INTERLEAVED GOLF AND DART TRIALS

Golf putting and dart throwing both generate U-shaped trends in runs z score with hit rate, but golf putting is much streakier than dart throwing. The difference in overall streakiness may be due to subtle differences in the designs of the respective studies. One way in which the golf putting and dart studies differed was in the pacing of trials. In the putting experiment, it was possible to continuously supply subjects with golf balls so that there was no break in activity; in the dart-throwing study, it was not

possible for the experimenter to continuously supply the subjects with darts without incurring the danger of being hit. Subjects in the dart study threw 3 to 5 darts and then collected them for another round. There was invariably a 5- to 10-sec delay between sets. Such delays render outcome sequences more Bernoulli-like in signal detection (Gilden & Gray Wilson, 1995) and may have caused a general suppression of streaks in dart throwing.

Time delay in the exercise of motor skills is relevant not only for understanding streaks in our dart-throwing experiment, but also for explaining the negative results reported by Gilovich et al. (1985) in the context of free-throw shooting. Gilovich et al. did not report the pacing of their free-throw study, and they may not have recognized the importance of allowing trials to be self-paced. We have assessed the importance of interruptions in motor activity by mixing golf trials with dart trials. In this design, dart throwing serves to interrupt golf putting and vice versa. We conjectured that interruptions would make both tasks resemble a Bernoulli process.

Method

Subjects. Four subjects were recruited by advertisement. They were paid \$5 per session.

Stimuli. The equipment was the same as that used in the earlier studies.

Design and Procedure. The subjects alternated activities in sets of 5 trials to accommodate the retrieval of golf balls and darts. A block consisted of 150 trials of each activity, and each subject completed 10 blocks. Half of the subjects began with darts, and half began with golf. Prior to the collection of data, we calibrated each subject's scores to achieve hit rates of about 0.5 in both activities.

Results and Discussion

Separate runs analyses were performed on the golf and darts parts of the sequences, as well as on the combined mixture. The mean runs z scores for the golf component, darts component, and mixture were .027, .01, and -0.26 , respectively. The isolated components were indistinguishable from the output of a Bernoulli process, but the combined activity, which is realistically what the subject experienced, was moderately streaky [$t(39) = -1.8, p < .05$]. We have checked that this result is not an artifact of the two games' being played at different hit rates. Our calibrations proved to be quite stable, and the resultant distribution of hit rates was contained in the interval $[.4-.65]$.

The results of this experiment may point the way to an ecologically sensible assessment of basketball. While a player is not shooting, he/she may be executing another activity (rebounding, passing, or whatever) with the élan that would be termed "in the zone" if it were shooting activity. Statistical analyses that concentrate on shooting activity may misrepresent the structure of the sport insofar as shooting opportunities are intermittent even when the action is continuous.

STREAK CAUSATION

There is some aspect of skilled performance that causes hits and misses to cluster together in specific regimes of task difficulty. In this section, we investigate several mod-

els of sequence structure in an effort to better understand what might be causing streaks to occur.

Learning is a prime example of hit-rate nonstationarity that could make performance appear to be streaky. A secular increase of hit rate over a block of trials could result in a deficit of runs relative to the expectation of a stationary Bernoulli process operating at the average hit rate. In this case, streaks would be an artifact of a process that might well be expected to occur and would be of little interest as a phenomenon in itself.

We have evaluated the role of learning in the second golf study (which showed the clearest evidence for streakiness) by constructing explicit learning curves that relate hit rate to trial number. While virtually any monotonic relationship between trial number and hit rate could serve as a model of learning, we have considered a representative and fairly exhaustive set of relations by allowing the hit rate to be a power law of trial number:

$$\text{hit rate} \propto (\text{trial number})^\beta, \quad \beta > 0.$$

This set includes all relations that are everywhere concave or convex.

We evaluated the importance of learning by partialing out the power-law models of hit rate from the outcome serial correlation in the observed data sequences. To the extent that a model describes the process whereby serial correlation arises, the part correlation will be smaller than the serial correlation. In the limit that the outcome sequence has no stochastic component and the learning model is an exact description, the part correlation vanishes. Let r_{12} represent the serial correlation (the correlation between the data sequence and itself lagged by one trial), r_{1M} the correlation between the sequence and the learning model, and r_{2M} the correlation between the sequence lagged by one trial and the learning model. Then the part correlation is

$$r_{\text{part}} = \frac{r_{12} - r_{1M}r_{2M}}{\sqrt{1 - r_{2M}^2}}.$$

If learning is occurring, then both r_{1M} and $r_{2M} > 0$ and $r_{\text{part}} < r_{12}$.

For a given sequence of outcomes, there is generally some value of β that minimizes r_{part} . Using standard numerical techniques, we have computed the minimum r_{part} for all sequences in the second golf study that had positive sequential dependency. The magnitude of the difference between the serial and part correlations is an index of the importance of learning in outcome clustering.

Gilden and Gray Wilson (1995) argued that streaks in signal detection were unrelated to learning and were due, rather, to oscillations in hit rate. We conjectured that a wave model would apply here as well. In wave models, the hit rate is conceived to vary as

$$\text{hit rate} \propto \sin(2\pi k/L + \theta),$$

where k is the trial number, L is the wave period (measured as number of trials), and θ is the phase. A wave in hit rate can easily produce positive sequential dependency; hits

congregate at the crests, and misses congregate in the troughs. Here both L and θ entered as free parameters in finding the minimum part correlation for each sequence.

The results from these analyses were straightforward. The part correlations for every sequence were smaller in the wave model. On average, the serial correlation for golf sequences was $r_{12} = .060$, while the part correlations were $r_{\text{part}}(\text{wave}) = .037$ and $r_{\text{part}}(\text{learning}) = .053$. (Although the magnitude of the serial correlation may not appear to be large, serial correlations on the order of 0.10 correspond to runs z scores more negative than -2.0 .) Relative to wave modulation, learning appears to be a rather small effect in explaining hit-rate structure.

Gilden and Gray Wilson (1995) considered a larger class of models that included a second-order Markov process and a stochastic version of a two-state wave model. The Markov model explicitly incorporates the idea that present outcome is correlated with earlier outcomes—"success breeds success." The stochastic two-state wave model (referred to as the intermittent effort model) simulates the fluctuations in performance that arise when the operator's attention drifts to and from the task. This model treats the hit rate as moving between two states dependent upon a transition probability. Stochastic occupation of states generates positive sequential dependency for the same reason the wave model does. Hits occur preferentially in the high-hit-rate state and misses in the low-hit-rate state. The primary difference between the wave and intermittent-effort models is one of determinism. In the wave model, selection of the phase, amplitude, and period fixes the hit rate on all trials.

Assessment of the Markov and intermittent-effort models requires considerably more effort than calculation of the part correlations. These models are inherently stochastic and require Monte-Carlo simulation. In this technique, a given hit-rate model prescribes a rule for generating pseudodata, sequences of binary digits that represent algorithmic performance. Gilden and Gray Wilson (1995) measured the local similarity between pseudo and real data sequences in terms of the probability of encountering m hits on n consecutive trials. In our signal detection studies, we found that the wave model generated pseudosequences that had the greatest resemblance to the data. We have repeated the same analyses on the golf data with the same result. Wave modulation of hit rate with a period of about 20 trials (20–50 sec given the pacing of trials) and an amplitude of 0.2 produces sequences that resemble golf data both globally (the runs z score) and locally (probability of m hits on n consecutive trials).

GENERAL DISCUSSION

Our finding that streaky performance in motor skills exists in laboratory studies does not contradict the main arguments of Gilovich et al. (1985) insofar as their work concerned the cognitive illusion of streaks in the full context of a basketball game. The work presented here does challenge the general claim, implicit in Gilovich et al., that streaks are always an illusion. To the contrary, we have found evidence not only for streaky performance, but also for a U-shaped function relating streak magnitude and hit rate. The difference between our results and those of

Gilovich et al. (1985) may be due to the pacing of trials. Delay between trials does tend to make outcomes independent. We have observed this in our first dart-throwing study, in the mixed golf and dart study, and generally in signal detection (Gilden & Gray Wilson, 1995).

We have also attempted to characterize the form of hit-rate nonstationarity that is associated with streaky performance. Two deterministic models (learning and wave modulation) and two stochastic models (Markov and intermittent effort) of hit rate have been considered. Wave modulation of hit rate provided the best description of observed sequence structure within this group of models. This result is consistent with our analysis of streaks in signal detection tasks (Gilden & Gray Wilson, 1995), and may point to a common etiology.

Our work is hardly the first to suggest that there are oscillations in performance. Circadian rhythms, for example, modulate performance quite generally (Hockey, 1986, reviews this field), but on time scales measured in hours. On the short time scales, 10–100 sec, implied by our models of outcome sequences, there is evidence for fluctuations in threshold that dates back to the beginning of psychophysical research (see Guilford, 1927, for a review). Guilford's account of threshold variability centered on retinal adaptation and eye movements, constructs that have no clear relevance to the exercise of motor skills. Threshold effects and adaptation of motor neurons may play a role in streak formation, although development of such a theory will require a more complete understanding of what is involved in skilled action.

Wave-like structures have not been reported in the domains of vigilance and controlled attention (see reviews by Davies & Parasuraman, 1982; Parasuraman, 1986). This may be due to several factors. Our inferences rely upon correlation minimization and Monte-Carlo simulation, numerical methods that are not common in the treatment of psychological data. In addition, our experimental design permits particularly sensitive tests of hit-rate models. Finally, we stress that the attentional demands made by skilled action are difficult to evaluate. Golf and darts implicate both controlled and automatic processes, and it is part of the lore of streak shooting that it occurs on those occasions when performance feels effortless.

REFERENCES

- CSIKSZENTMIHALYI, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.
- CSIKSZENTMIHALYI, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.
- CSIKSZENTMIHALYI, M., & CSIKSZENTMIHALYI, I. S. (Eds.) (1988). *Optimal experience: Psychological studies of flow in consciousness*. New York: Cambridge University Press.
- DAVIES, D. R., & PARASURAMAN, R. (1982). *The psychology of vigilance*. New York: Academic Press.
- GILDEN, D. L., & GRAY WILSON, S. A. (1995). On the nature of streaks in signal detection. *Cognitive Psychology*, **28**, 17-64.
- GILOVICH, T., VALLONE, R., & TVERSKY, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, **17**, 295-314.
- GUILFORD, J. P. (1927). Fluctuations of attention with weak visual stimuli. *American Journal of Psychology*, **38**, 534-583.
- HAYS, W. L. (1988). *Statistics* (4th ed.). New York: Holt, Rinehart & Winston.
- HOCKEY, G. R. J. (1986). *Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms*. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 44-1 to 44-49). New York: Wiley.
- KUBOVY, M., & GILDEN, D. L. (1989). Apparent randomness is not always the complement of apparent order. In G. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 115-127). Washington, DC: American Psychological Association.
- LARKEY, P. D., SMITH, R. A., & KADANE, J. B. (1989). It's okay to believe in the "hot hand." *Chance*, **2**, 22-30.
- PARASURAMAN, R. (1986). Vigilance, monitoring, and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, 43-1 to 43-39). New York: Wiley.
- 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. *Science*, **185**, 1124-1131.

(Manuscript received April 26, 1994;
revision accepted for publication November 27, 1994.)