Title
Randomness in binary sequences: Conceptualizing and connecting two recentdevelopments

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Author
Reimers, Stian

Publication Date
2017

Peer reviewed

Randomness in binary sequences: Conceptualizing and connecting two recent developments

Stian Reimers (stian.reimers@city.ac.uk)
Department of Psychology, City, University of London
Northampton Square, London EC1V 0HB, United Kingdom

Abstract

Recent theoretical research has shown that the assumptions that both laypeople and researchers make about random sequences can be erroneous. One strand of research showed that the probability of non-occurrence of streaks of repeated outcomes (e.g., HHHHHH) is much higher than that for a more irregular sequence (e.g., HTTHTH) in short series of coin flips. This tallies with human judgments of their likelihood of occurrence, which have conventionally been characterized as inaccurate and heuristic-driven. Another strand of research has shown that patterns of hits and misses in games like basketball, traditionally seen as evidence for the absence of a hot-hand effect, actually support the presence of the effect. I argue that a useful way of conceptualizing these two distinct phenomena is in terms of the distribution of different sequences of outcomes over time: Specifically, that streaks of a repeated outcome cluster whereas less regular patterns are more evenly distributed.

Keywords: randomness; rationality; hot hand fallacy, gambler's fallacy.

Introduction

One of the more important things that organisms must do to prosper is to identify, extract, and act on patterns in the environment. At a perceptual level, detecting potential threats in a noisy and ambiguous environment is crucial for survival. At a higher level, the ability to detect patterns of events over time or space, such as the presence of absence of prey in different locations at times, the changes in temperature or weather, and so on, allows an organism to predict the future state of the world, and adapt behavior accordingly. In a more contemporary environment, anyone able to detect behavioral patterns in markets, organizations or individuals would be able to exploit that knowledge to their benefit.

In order to detect patterns, an organism has to separate signal from noise. As such, one would expect organisms to accurately represent the absence of a signal, that is, randomness. A poor representation of what random patterns look like would make it harder to spot the times when patterns contain information.

As such, it is surprising that across a wide range of research procedures, people are systematically poor at representing randomness (for reviews see, Nickerson, 2002, 2004; Bar-Hillel & Wagenaar, 1991; Falk & Konold, 1997; Rapaport & Budescu, 1992). For example, people underestimate the frequency of 'streaks' or 'runs' of a particular outcome (such as getting five heads in a row when flipping a coin repeatedly), and treat such streaks when they appear as evidence for non-randomness. Related, people rate sequences of binary outcomes containing negative serial dependency (that is, an alternation rate between outcomes of greater than .5), as being more random than truly random sequences.

One reason why people may be poor is that many properties of random sequences are counterintuitive. For example, relative wait times for different sequences of binary outcomes violate transitivity (see, e.g., Nickerson, 2007).

In this paper I focus on sequences of binary, Bernoulli i.i.d. events such as coin flips (which could come down H or T), and behaviors which may be modelled by them, such as basketball shots (which could come down as a hit – X – or a miss – O).

Representativeness and probability of occurrence

One of the most influential studies to demonstrate an apparent bias in perception of randomness was that of Kahneman and Tversky (1972). In their studies, they asked participants about the relative frequency of occurrence of different birth orders of girls (G) and boys (B) in families with six children in a hypothetical city. They found that participants estimated that there would be many fewer examples of a precise sequence of BGBBBB relative to a precise sequence of GBGBBG. (Of course all precise sequences of birth orders are equiprobable). To examine whether this finding was a result of just the relative frequency of B and G, Kahneman and Tversky also compared estimates of the relative frequency of BBBGGG and GBGBBG, finding that the former was seen as significantly less probable than the latter. Thus, both the relative frequency of outcomes, and the order in which outcomes occur appear to be important in judging the probability of occurrence.

Traditionally, findings of this nature have been explained in terms of heuristics and biases, specifically a misapplication of a representativeness heuristic (but see, e.g., Gigerenzer, 1996; Ayton & Fischer, 2004): People believe that the properties of short sequences of random outcomes should be representative of those seen in longer sequences (e.g., equal proportions of outcomes, an absence of structure or compressibility), and sequences that share those properties are deemed more probable.

However, recently Hahn and Warren (2009) observed that in situations where one looks for patterns of outcomes in a finite sequence of, for example, coin flips, different sequences have different probabilities of occurrence.

To give a concrete example (used by Hahn and Warren, 2009), compare the probability of *non*-occurrence of a HHHH vs HHHT in a sequence of 20 coin flips. The streak of a repeated outcome (HHHH) is around twice as likely not to occur relative to HHHT. The argument made by the authors is that if people use previous experience of merely the occurrence (at least once) or absence of a particular string to judge the probability of occurrence in the future, then they would be quite accurate to say that HHHH would be less likely to occur in a sequence of 20 coin flips than HHHT.

This was also extended to account for the gambler's fallacy: If experience dictates that HHHH is less likely to occur than HHHT, then an individual who sees HHH and is asked to be on whether the next observation is H or T, would with some justification bet on T.

Although there is ongoing discussion about the extent to which or circumstances under which Hahn and Warren's theory predicts judgments (Reimers, Donkin & Le Pelley, 2017), the observation that different strings of outcomes have different probabilities has meant that researchers have needed to reconsider what normative baselines for randomness judgement should be, and potentially turn what previously appeared to be a clear bias into a slight misapplication of a genuine property of the environment.

The hot-hand-fallacy fallacy

A second challenge to researchers' assumptions about normative baselines has been seen with the hot hand effect.

The hot hand effect is a phenomenon – accepted as self-evidence by many sports participants and spectators – that players go through periods when their performance varies consistently over time, having streaks when they are 'hot', and during that period of time their performance is consistently better than usual, as measured by, for example, the proportion of baskets or putts they manage to sink. If the hot hand were real, it would mean that probability of success had positive autocorrelation: Following a streak of hits, a person would be more likely to score another hit.

Despite popular belief in the hot hand phenomenon, the effect has until recently been seen as a fallacy. Gilovich, Valone and Tversky (1985) examined the performance of professional and amateur basketball players, and argued that there was no evidence for a hot hand effect. They operationalized a hot hand effect in basketball shooting as a difference between the probability of getting a hit (scoring from a free throw) after a streak of k consecutive hits (X) and the probability of getting a hit after a streak of k successive misses (O), for example, $p(X|XXX) > p(X|OOO)$. The logic, which appears superficially entirely reasonable, was that if the probabilities of a hit after k hits and a hit after k misses were identical in a well-power study, then that provided evidence for the absence of a hot-hand effect. Across several studies, they found no difference in probabilities, so concluded that the hot-hand effect was a fallacy.

Recently this conclusion has been challenged. Miller and Sanjurjo (2016) note problems with measures traditionally used to support the absence of a hot hand effect (see Rinott and Bar-Hillel, 2015 for less technical overview of an earlier version). Specifically, they prove that if one were to calculate the strength of the hot hand effect for players individually along the lines of calculating $p(X|XXX) / [p(X|XXX) + p(X|OOO)]$, and then take the average across individual, that average would be less than .50. So if a well-powered study shows a mean proportion of around .50, then rather than being evidence against a hot hand effect, it is in fact substantial evidence for such an effect.

Miller and Sanjurjo (2016) prove the counterintuitive finding that that in any finite binary sequence, the mean proportion of streaks of length k that are followed by a repetition of the same outcome is on average lower than the proportion of streaks of length k that are followed by the opposite outcome. They note that for $k = 1$, the effect is entirely driven by a sampling-without-replacement effect, such that in, say, a short sequence of coin flips where the number of heads and tails is expected to be identical, choosing to look at outcomes following a H removes a H from the sample, meaning that the probability of all other observations, including the next one being a T is slightly greater than .5.

More relevant for this discussion is the effect where $k > 1$. Here, Miller and Sanjurjo note that the effect is driven much more by the extent to which sequences of outcomes can overlap with each other (or show autocorrelation, in the terminology of Guibas & Odlyzko, 1981). They note that some sequences of outcomes can overlap with themselves: For example the sequence HHH can partially overlap with itself such that in a series of five coin flips, it is possible to observe three overlapping instances: HHHHH; conversely, the sequence HTT cannot overlap at all, and so can only occur once in a series of five coin flips. They note that because overall the expected number of occurrence of HHH and HTT must be identical, HTT must be observed in a greater number of series of five coin flips to compensate for the fact that HHH can occur multiple times within a single series. As such, they prove that

Variance of occurrences in short sequences

Here, in contrast to Miller and Sanjurjo's (2016) formal proof, a stochastic approach to this issue is taken, in part to make the relationship between the findings of Hahn and Warren (2009) and Miller and Sanjurjo (2016), and in part to attempt to show how the varying distribution of observations of different sequences of outcomes in a longer series of binary outcomes can account for both findings.

This is not the first attempt to relate these two phenomena. In recent iterations of their working paper, Miller and Sanjurjo have attempted to account for the gambler's fallacy as well as the hot hand fallacy, by assuming a degree of insensitivity to sample size. Sun and

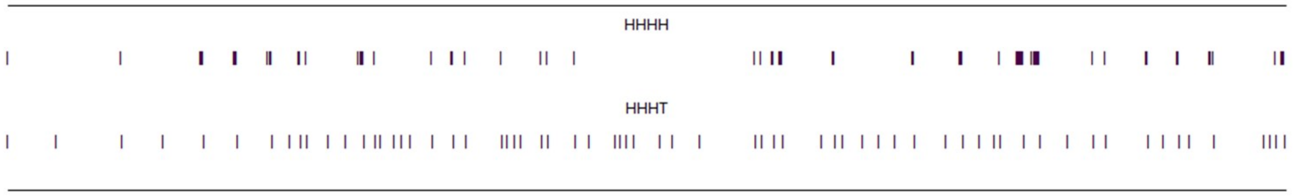

Figure 1: Raster plot of the occurrence of strings of HHHH or HHHT across 1,000 simulated coin flips.The horizontal dimension gives the flip from the first on the far left to to the last on the far right.

Wang (2010) note that different forms of waiting time for sequences of outcomes vary differently with outcome. Thus, the mean inter-observation gap is the same for all sequences of a single length, whereas the expected waiting time from first flip of a coin is much greater for some sequences of outcome (such as HHHH) than others (HHHT), and that the variances in these two forms of waiting time vary substantially.

The argument presented here is based on the observation that the variance of the number of trials between observations of a given sequence of outcomes varies. Specifically, the observations made by Hahn and Warren, and those made by Miller and Sanjurjo are both consequences of the same property of random sequences, specifically that within any finite sequence of equiprobable binary outcomes, the distribution of frequency-of-occurrence for 'streaks' (i.e. repetitions of the same outcome, like HHHH) is much wider than that for non-streaks (like HHHT).

To compare the distribution of two sequences of outcome HHHH and HHHT, across 1,000 simulated coin flips, see Figure 1. As both Hahn and Warren (2009) and Miller and Sanjurjo (2016) note, although the total number of occurrences of HHHH and HHHT is approximately equal, HHHH tend to cluster more than HHHT, with several overlapping occurrences together, and then large gaps between them. One way of explaining this it is that we know that overall frequency of HHHH and HHHT must be on average identical. However, immediately after flipping HHHH, there is a 50% chance of flipping another head, giving another instance of HHHH, and then a 50% chance of another, and so on. This leads to clusters of consecutive overlapping instances of HHHH. Conversely, after flipping HHHT, it takes a minimum of four more flips to get HHHT again. This means that HHHT cannot cluster in the same way.

The consequence is that for shorter sequences of, say, 100 random binary outcomes, the frequency of HHHT will be fairly consistent, whereas the frequency of HHHH will be much more variable. This can be seen in Figure 2, in a simulation of 10,000 runs of 100 coin flips. Here, the string HHHT appears between 3 and 9 times on 95% of runs of 100 flips. HHHH only appears between 3 and 9 times on 67% of runs.

Hahn and Warren's theory explains the fact that people seem to think HHHH is less likely to occur than HHHT, by looking at the difference in the probability of non-occurrence of a string (or conversely the probability of its occurring at least once). Although they use shorter runs for their examples, the same pattern is observed: In Figure 2, the string HHHH is much more likely not to occur than HHHT is. In fact, although it is hard to see from the graph, the probability of HHHH's non-occurrence is around 100 times that of HHHT. This is – of course – a consequence of the fact that the mean of the frequency-of-occurrence distribution for HHHH is the same as that for HHHT, but the variance is much greater. Hahn and Warren suggest that when making judgments, people, whose experience is limited to short runs of outcomes, might attend to whether a string occurs or not, but not attend to the number of times it occurred. This means that they will see HHHT occurring in a lot more runs than they will HHHH, and will rate it more probable.

Conventional analysis of Gilovich et al.'s hot hand data used the logic that in the absence of a hot hand effect, the average proportion of players' shooting successes would be the same following three successes as following three failures. Miller and Sanjurjo note that this is not the case. The observation I make here is that this is a direct consequence of the distribution of frequency-of-occurrence in 100 binary outcomes being much wider for streaks than non-streaks is that the proportion of XXXX from {XXXX, XXXO} (or, by symmetry {XXXX, OOOX}) is less than .5.

To give a concrete example, if every day your grocer randomly gives you either 3, 4 or 5 apples, and either 2, 3, 4, 5, or 6 oranges, and each day you work out what proportion of the fruit you were given is apples, you will find that, averaging across many days, the proportion of apples is greater than .5, even though the total number of apples and oranges you receive is on average identical.

(If this is not obvious, consider the case where the grocer always gives you 4 apples, and also randomly gives you either 0 or 8 oranges. Half the time you leave with a bag that contains 100% apples; half the time you leave with a bag that contains 33% apples, so overall, the proportion of fruit in your bag that is apples averages 67%. However, the overall number of apples and oranges you receive will be the same.)

Thus, in general, if one draws a sample from two distributions which have the same mean but different variances, and then looks at the proportion of the combined outcome that comes from each distribution, the expected

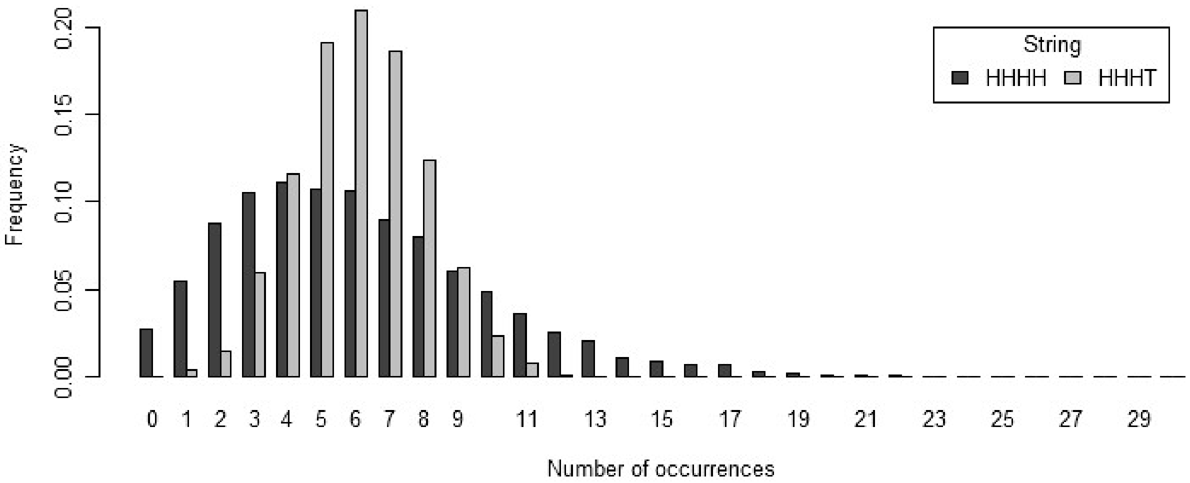

Figure 2: Distribution of frequency-of-occurrence for two different strings of outcomes in 10,000 simulated sequences of 100 coin flips

proportion from the lower variance distribution will be greater than that for the higher variance distribution.

This phenomenon can be seen more generally in Figure 3, which takes a normal approximation of the frequency-of-occurrence distributions shown in Figure 2, with equal means, capping at 0, and varying the SD of the two frequencies-of-occurrence. The color indicates the mean proportion of outcome y, averaged across 50,000 simulated trials of each of 100 random binary outcomes. Where SDs are equal, then of course p(y) = p(x) = .5. Where SD(y) > SD(x), p(y) < p(x), and vice versa. A circle indicates the approximate point where SD(x) = SD(HHHH) and SD(y) = SD(HHHT).

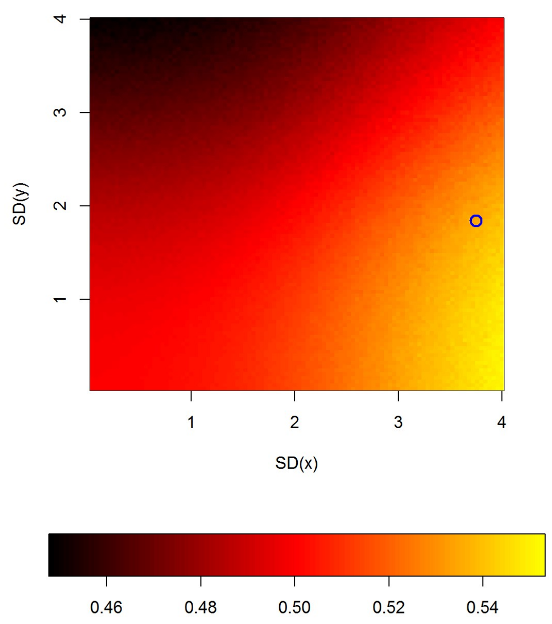

Figure 3: Simulated samples drawn from distributions with a common mean, Color indicates the mean proportion of the sum of the two samples that comes from the y sample

Replacing HHHH and HHHT with XXXX and OOOX, it is clear that, as Miller and Sanjurjo (2016) note, it is not correct to assume that, in the absence of a hot hand effect, the expected proportion of successes following *k* successes, averaged across a large set of players, should be .5. Rather, it is significantly lower, as a direct consequence of the distribution of frequency-of-occurrence for XXXX being broader than that for OOOX.

Conclusion

The argument presented here is that both Hahn and Warren (2009) and Miller and Sanjurjo's (2016) findings can be explained the same way: In sequences of random binary outcomes, streaks of the same outcome (whether heads, HHHH, or successes, XXXX) cluster more than non-streaks (HHHT, OOOX); this leads to a broader distribution of frequency-of-occurrence of streaks in finite sequences of random binary events relative to non-streaks. This both increases the chance of the non-occurrence of a streak (which H&W argue makes people think justifiably that HHHH is less likely to occur than HHHT and other non-streaks) and reduces the average proportion of XXXX among observations of {XXXX and OOOX} (which Miller and Sanjurjo convincingly argue means that evidence for a hot hand effect has been overlooked).

There are potentially interesting implications from these observations for the kinds of cognitive representation that would mediate the biases seen here. For example, an agent that counted the total number of occurrences of different strings of outcomes would see that the number of occurrences of, say, HHHH and HHHT were identical, so should rate them as equally probable. An agent that discarded all information about the frequency of occurrence of a string and recorded only whether or not it was observed (at least once) in a particular set of connected outcomes would of course perceive HHHT as more common than HHHH. Similarly, an agent that, rather than counting the number of occurrences of a string, instead encoded only the relative frequency of different strings, as a proportion of the total number of observations across occasions, would also conclude that HHHT was more frequently observed than HHHH.

(Of course the overlapping of streaks described above may account for the biases seen here in more superficial ways. Chater (2014) argues that cognitive segmentation processes may differentially mask the frequency of occurrence of different strings. For example, a sequence of TTHHHHHHTT might be parsimoniously chunked as two tails – six heads – two tails, underplaying the three overlapping occurrences of HHHH within the sequence.)

Overall it seems clear that an examination of the distribution of frequency-of-occurrence for different strings of binary outcomes, allows one to create a parsimonious and intuitive account for two important recent theoretical observations, both of which have implications for the study of rationality.

Acknowledgments

Thanks to Peter Ayton, Adam Harris, and Mike Le Pelley, for helpful discussions on this topic.

References

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? Memory & Cognition, 32, 1369–78.

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

Chater, N. (2014). Cognitive science as an interface between rational and mechanistic explanation. *Topics in Cognitive Science, 2*, 331-337.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*, 592–596.

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

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. Psychological Review, 104, 301-318.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: Why three heads are better than four. *Psychological Review, 116,* 454-461.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430-454.

Miller, J. B. & Sanjurjo, A., (2016) Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers. IGIER Working Paper No. 552. Retrieved February 2, 2017 from: https://ssrn.com/abstract=262735. doi: 10.2139/ssrn.2627354

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109,* 330–357. doi:10.1037//0033-295X.109.2.330

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Mahwah, NJ: Erlbaum

Nickerson, R. S. (2007). Penney Ante: Counterintuitive probabilities in coin tossing. *The UMAP Journal, 28,* 503–532

Rapoport, A., & Budescu, D. V. (1997). Randomization in individual choice behavior. *Psychological Review, 104,* 603–617.

Reimers, S., Donkin, C., & Le Pelley, M. E., (2017). Perceptions of randomness in binary sequences: Normative, heuristic, or both? *Manuscript in preparation.*

Rinott, Y., & Bar-Hillel, M. (2015). Comments on a 'hot hand' paper by Miller and Sanjurjo (2015). http://ssrn.com/abstract=2642450. doi:10.2139/ssrn.2642450

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