Relatively random: Context effects on perceived randomness and predicted outcomes

William J. Matthews
University of Essex

Correspondence Address:

William J. Matthews
Dept. of Psychology
University of Essex
Colchester
CO4 3SQ
United Kingdom

Tel.: +44 1206 873818
Fax: +44 1206 873801
E-mail: will@essex.ac.uk

Published as: Matthews, W.J. (2013). Relatively random: Context effects on perceived randomness and predicted outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1642-1648.

Abstract

This paper concerns the effect of context on people's judgments about sequences of chance outcomes. In Experiment 1, participants judged whether sequences were produced by "random", mechanical processes (such as a roulette wheel) or skilled human action (such as basketball shots). Sequences with lower alternation rates were judged more likely to result from human action. However, this effect was highly context-dependent: A moderate alternation rate was judged more likely to indicate a random physical process when encountered amongst sequences with lower alternation rates than when embedded among sequences with higher alternation rates. In Experiment 2, predictions regarding the next outcome following a streak showed the same effect: A streak of three at the end of the sequence was judged less likely to continue by participants who had encountered shorter terminal streaks in previous trials than by those who had encountered longer ones. These contrast effects (a) help to explain variability in the types of sequences that are judged to be random and that elicit the gambler's fallacy, and urge caution about attempts to establish universal parameterizations of these effects, (b) are congruent with theories of sequence judgment which emphasize the importance of people's actual experiences with sequences of different kinds, (c) provide a link between models of sequence judgment and broader accounts of psychophysical/economic judgment, and (d) may offer new insight into individual differences in randomness judgments and sequence predictions.

Keywords: Context effects; perceived randomness; gambler's fallacy; hot hand

Relatively random: Context effects on perceived randomness and predicted outcomes

Probabilistic sequences of binary alternatives are ubiquitous in both the laboratory and the outside world: Stimuli are "signal" or "noise", coin flips are "heads" or "tails", and papers are rejected or accepted. How people represent, interpret, and produce such sequences is a central aspect of cognition that has been extensively researched (see Oskarsson, Van Boven, McClelland, and Hastie, 2009, for a review) and has important practical implications (e.g., Kwan, Wojcik, Miron-shatz, Votruba, & Olivola, 2012).

Two aspects of sequence judgment have received particular attention. The first concerns perceptions of the generating process: How do people decide what kind of mechanism produced a given sequence of outcomes? Studies in which participants judge the randomness or likely source of sequences have shown that a key determinant of subjective randomness is the alternation rate (AR) – a measure of how often the sequence switches from one outcome to the other. (For example, the sequence @@@##@ alternates twice out of a possible five times, giving an AR of 0.4.) Sequences judged most random typically have alternation rates greater than 0.5. For example, Ayton and Fischer (2004) had participants judge whether sequences of @ and # symbols represented the outcomes of a random mechanical process (e.g., the heads and tails produced by flipping a coin) or skilled, intentional human action (e.g., the successful/unsuccessful shots of a professional basketball player). Increasing the AR from 0.2 to 0.8 produced a steady increase in judgements that the sequence was produced by the mechanical process.

A second area of research concerns predictions of future outcomes. The most well-known finding is that a streak of one outcome leads people to predict the opposite outcome on the next trial, even when the outcomes are strictly independent. This has been found both when people attempt to predict stochastic outcomes one after the other and when they are given a summary description of the sequence (e.g., Ayton & Fischer, 2004; Burns & Corpus, 2004). It also occurs when people attempt to postdict the event that preceded a streak (Matthews, 2010) and arises in real financial decisions (e.g., Sundali & Croson, 2006). This belief in negative recency is usually labelled the Gambler's Fallacy (GF).

The inverse tendency is also widely documented. That is, in some situations a streak of one outcome leads to the prediction that the next trial will follow the streak (see Oskarsson et al., 2009). This positive recency is labelled belief in the hot hand (HH) after the demonstration by Gilovich, Vallone, & Tversky (1985) that basketball fans expect a player who is on a scoring streak to have an elevated chance of success with his next shot, when in fact there is no such positive recency in players' performance (although some sports do show such hot and cold spells -- see e.g., Smith, 2003).

The most widespread theoretical account of these phenomena is the idea that judgements are made using a representativeness heuristic, such that people expect local regions of a random sequence to be representative of their prototype or schema for random processes (Kahneman & Tversky, 1972). For example, a streak of four heads in a row is judged unlikely because it violates the characteristic of random sequences that both outcomes occur equally often, leading to elevated expectations of "tails" for the next flip (the GF). The same idea has been invoked to explain belief in the hot hand: A run of one outcome is judged unrepresentative of a random process, leading people to conclude that successive outcomes are positively correlated (Gilovich et al., 1985). Alternative accounts include the suggestion that the GF reflects a generalization of expectations from the sampling-without-replacement that characterizes many physical processes, whereas belief in the hot hand arises from a generalization of the positive recency that characterizes many aspects of human performance (e.g., Ayton and Fischer, 2004). Other theorists have emphasized folk ideas about luck and randomness (see Oskarsson et al., 2009), analyses of the waiting times for particular sequences (e.g., Hahn & Warren, 2009), or subjective complexity (Falk & Konod, 1997).

A major challenge for theories of sequence judgment is that sequences with identical or similar properties can produce very different judgments. For example, Burns and Corpus (2004) found that the same sequence was judged less random and more likely to continue a streak when framed as a competitive human interaction than as the spins of a roulette wheel; likewise, instructing participants to construe outcomes as "like the flips of a coin" increased the prediction that a streak will end relative to a condition where they were given no specific instructions about the origin of the sequence (Boynton, 2003), and telling participants that sequences represented random outcomes led them to remember streaks as shorter than when the same sequences had been described as being produced by a complex algorithm (Olivola & Oppenheimer, 2008). Similarly, different studies find different critical values of sequence parameters. For example, the streak length required to elicit a significant gambler's fallacy differs between studies (e.g., Croson and Sundali, 2005; Barron & Leider, 2010). Finally, there are marked individual differences in the tendency to show both the GF and HH (e.g., Sundali & Croson, 2006).

There are doubtless many reasons for this variability. The current work examines the contribution of one potentially important factor, namely the context established by other recently encountered sequences. The judgments made in the foregoing studies did not occur in isolation; the sequences were experienced in the context of other items presented during the experiment and encountered outside the laboratory, and this context may exert a marked influence on the judgment of a given sequence. Specifically, judgments can assimilate towards or contrast away from the other items in a stimulus ensemble. Assimilation can arise for several reasons. For example, the perception of the target may be a weighted average of that stimulus and the other stimuli in the set, each item may be judged relative to the previous one, or the participant may simply tend to repeat responses (see Matthews & Stewart, 2009, for a discussion). Contrast can arise when judgment of a given item is based on its distance from the mean of the set or its rank position (e.g., Helson, 1964; Parducci, 1965), or when people search for differences between the target and the preceding item (e.g., Damisch, Mussweiler, & Plessner, 2006). Assimilation and contrast effects have been found in various domains, including perceptual, social, and economic judgments, although they are not ubiquitous and are not always substantial (see e.g. Stewart, Chater, Stott, & Reimers, 2003).

Whether perceived randomness and outcome prediction are context-dependent is important because theoretical development requires that we quantify the sequence properties that determine judgments. For example, Carlson and Shu (2007) have argued for a "rule of three", such that "the third repeat event in a sequence is pivotal to the subjective belief that a streak is occurring" (p. 113). Similarly, it has been posited that alternation rates of 0.6 are regarded as "most random" (e.g., Oskarsson et al., 2009). It is important to establish the stability of such estimates to changes in context, partly to avoid incorporating erroneous parameter values into formal models, but also because the sources of parameter variation can guide the interpretation of the parameter estimates themselves. In addition, some theorists assume that sequence perceptions and predictions are a consequence of the kinds of sequences that humans typically experience. For example, Wilke and Barrett (2009) have argued that belief in the hot hand is an "evolved psychological default" based on foraging for clumpy resources, and can be reduced by experience with genuinely independent events. Central to such theorizing is the idea that perceptions of a given sequence can shift as a consequence of experience with other (unrelated) sequences – in other words, that they are context dependent. Finally, identifying and understanding contextual effects on judgments of randomness and sequence predictions may contribute to integrated theoretical development. Theories of contextual effects have been developed and applied across a range of perceptual, social, and economic domains (see below), and exploring the effect of context on sequence judgments may offer a step towards more general theoretical unification.

Experiment 1

Experiment 1 used a procedure similar to Experiment 2 of Ayton and Fischer (2004). Participants were presented with sequences of outcomes and judged whether each sequence was more likely to have been produced by a "random" physical process (coin flips, roulette wheels, dice) or an intentional human action where positive recency may be the (perceived) norm (basketball, soccer, tennis). Each sequence comprised 21 outcomes coded as @ (n=11) and # (n=10), in random order, with the alternation rate chosen to be either low (0.15), medium (0.5), or high (0.85). All participants saw sequences with a medium alternation rate, but for some these were intermixed with low-AR sequences whilst for others they were intermixed with high-AR sequences. The key question was whether judgments about the process producing the medium-AR sequences depended on the context established by the other sequences in the experiment.

Method

Participants. Participants took part on-line in exchange for a small payment and were recruited via Amazon Mechanical Turk using the Crowdflower crowdsourcing platform (www.crowdflower.com). The final sample comprised 165 participants (97 male, ages 18-74 years, M= 31.9, SD = 12.0)[1].

Design and Procedure. The task was described as follows:

"This study concerns your ability to recognize sequential patterns from various sources. Many processes produce a sequence of outcomes. For example, a coin may be tossed several times to produce a sequence of "heads" and "tails", or a basketball player may take many shots to produce a sequence of "baskets" and "misses".

[1] To ensure independent responses, only the first occurrence of a given ip address was used both within and between the two experiments, and a small number of participants were excluded for failing to complete the task (19 in Experiment 1; 3 in Experiment 2). Final sample sizes are after these exclusions.

On the following pages, you will be shown several sequences of outcomes. The outcomes are represented by the symbols @ and #. For example, one sequence might look like this:

@@#@@###@@@##@#@@###@

So, you will not know what each outcome is, but you will be able to see the pattern of the sequence. Your task will be to judge which of two processes is more likely to have produced each sequence."

The following three pages each presented 12 sequences for judgment; each sequence was preceded by the words "What produced this sequence?" and was followed by a six-point rating scale, with lower numbers indicating greater likelihood that the sequence was produced by a random physical process and higher numbers indicating that the sequence was produced by skilled/intentional human performance. Three pages presented different versions of this task. In one scenario, the choice was between a series of coin flips (heads vs. tails) and a series of basketball shots (score vs. miss); for another, the choice was between roulette wheel spins (red/black) and soccer shots (score/miss); in the third scenario, the choice was between dice rolls (lands on even number/lands on odd number) and tennis serves (in/out). A preamble at the top of each page explained the two options. The extremes of the 6-point rating scale were given verbal labels appropriate to the choice (e.g., "Definitely Coin" and "Definitely Basketball"). Responses were required for all questions before the participant could progress to the next page.

Each sequence had one of three alternation rates: 0.15 (low), 0.5 (medium), or 0.85 (high). The sequences were produced by randomly generating outcomes subject to the constraint that the final sequence had the desired AR. Sequences with the medium alternation rate are referred to as "target" sequences. A set of 6 target sequences was randomly intermixed with 6 low-AR sequences; the same 6 target sequences were also randomly intermixed with 6 high-AR sequences. This gave two sets of 12 sequences, one where the target sequences were embedded in a high-AR context and another where the same targets were embedded in a low-AR context. This process was repeated 3 times to give three "low-AR context" sequence sets and 3 "high-AR context" sequence sets. Participants in the low-AR context condition were presented with the three low-AR context sets (one for each of the coin-basketball, roulette-soccer, and dice-tennis scenarios); participants in the high-AR context condition were given the three high-context sets. Examples of the sequences are given in Table 1. Assignment of sequence sets to the different scenarios and order of scenario presentation were pseudorandom. The order of the sequences on each page was randomized for each participant. Eighty five participants were assigned to the low-AR context and 80 to the high-AR context.

Results and Discussion

For each participant and each scenario, mean judgments were calculated separately for the context sequences and for the target sequences. The results are shown in Figure 1.

We begin by examining the effect of alternation rate. In all three scenarios, independent-samples *t*-tests showed that the sequences with low alternation rate were judged more likely to have been produced by skilled human action than were the sequences with high alternation rate [for coin-basketball: $t(163) = 9.63$, $d = 1.51$ $p < .001$; for roulette-soccer, $t(163) = 7.37$, $d = 1.15$, $p < .001$; for dice-tennis, $t(163) = 7.19$, $d = 1.13$, $p < .001$]. Similarly, paired-samples *t*-tests showed that the medium-AR (target) sequences were judged more likely to be the result of skilled human action than the high-AR sequences [for coin-basketball, $t(79) = 6.64$, $d = 1.69$, p <.001; for roulette-soccer: $t(79) = 4.73$, $d = 1.16$, $p < .001$; for dice-tennis: $t(79) = 3.00$, $d = 1.05$, $p = .004$], but more likely to be the result of a random mechanical process than the low-AR sequences [for coin-basketball, $t(84) = 7.78$, $d = 1.57$, $p < .001$; for roulette-soccer: $t(84) = 6.04$, $d = 1.16$, $p < .001$; for dice-tennis: $t(84) = 8.40$, $d = 1.61$ $p < .001$]. This pattern replicates the findings of Ayton and Fisher (2004).

More importantly, judgments about the target sequences (which always had the same moderate alternation rate) depended upon the other sequences encountered in the experiment. For all three scenarios, target sequences were judged more likely to have been produced by skilled human performance in the high-AR context condition than in the low-AR context condition (for the coin-basketball scenario: $t(163) = 9.17$, $d = 1.44$, $p < .001$; for roulette-soccer, $t(163) = 5.66$, $d = 0.89$, $p < .001$; for dice-tennis[2], $t(147.5) = 6.37$, $d = 1.01$, $p < .001$). Thus, there was a marked contrast effect in judgments about the process responsible for generating the target sequences.

In short, sequences with rapid alternations between outcomes were judged more likely to have been produced by a random mechanical process (and less likely to have been

[2] Here and at certain points below the df are non-integer because a Welch correction has been applied because of unequal variances.

produced by skilled, intentional human action) than sequences with low alternation rates. However, judgments about a given sequence were strongly affected by the other sequences presented for judgment: If the context contained sequences with low alternation rates then sequences with a medium AR seem to have appeared less “streaky” and more “random” by comparison, and were correspondingly judged more likely to have been produced by a random mechanical process like a roulette wheel. On the other hand, if the context contained many sequences with high alternation rates (lots of short streaks) then the same moderate AR seems to have appeared “non-random” by comparison, and more likely to be the product of skilled human action. This contrast effect was substantial and there was no indication that judgments of the target sequences assimilated towards judgments of the context items. Experiment 2 extends these findings by examining predicted outcomes.

Experiment 2

In Experiment 2, participants predicted the next outcome in a sequence. The target sequences always ended with a streak of three, and context was manipulated between participants. For participants in the short-streak context condition, the target sequences were intermixed with sequences that ended in a streak of just one outcome. For those in the long-streak context condition, the target sequences were intermixed with sequences that ended in a streak of five. Two versions of the task were constructed: one group of participants was told that the sequences corresponded to the flips of a coin; another group was told that the sequences depicted the outcomes of basketball shots.

Method

Participants. Participants were recruited on-line and screened as before, giving a sample of 433 (252 male, 2 decline to indicate gender; ages 18-69 years, *M* = 29.9, *SD* = 10.7)[3].

Design and Procedure. Participants who completed the Coin version of the task were shown the following instructions:

[3] A larger sample was used than in Experiment 1 as a smaller effect was anticipated.

"On the next few pages, you will be asked to consider sequences of coin flips. The outcome of each flip is indicated by a letter, T for Tails and H for Heads. You will see a sequence of outcomes and be asked to indicate what you think the next outcome will be. There are no right or wrong answers -- we are interested in your gut feeling."

They then saw a series of 12 sequences. The sequences were presented one at a time on successive pages. Each sequence comprised 20 outcomes, 10 Heads and 10 Tails, that ended with a streak of length one (a single "H" or a single "T"), a streak of three ("H H H" or "T T T") or a streak of five ("H H H H H" or "T T T T T"). Each sequence ended with a prompt "_____?" followed by the words "What will happen next? Choose a number from 1 to 6, where larger numbers indicate greater likelihood that the next flip will be tails". Below this was a six-point response scale anchored at 1 ("Definitely Heads") and 6 ("Definitely Tails"). A response was required before participants could progress to the next sequence.

A terminal streak of length n constrains the last $n – 1$ outcomes (because the outcome before the streak must be of the opposite type). For each sequence with a given terminal streak length, the remaining 20-n-1 outcomes were randomly chosen, subject to the constraint that, over the whole sequence, there be 10 of each outcome and an alternation rate of 0.5.That is, the sequences were matched for their outcome probabilities and overall "streakiness".

Sets of 18 sequences were constructed. In each set, 6 sequences ended with a short streak (one outcome); 6 ended with a medium streak (3 outcomes); and 6 ended with a long streak (5 outcomes). (Streaks of heads and tails were used equally often for each streak length). The medium-streak items formed the target sequences. The 6 short-streak sequences were randomly paired with the target items to give a "short-streak context" set of 12 sequences (the same pairing was used across participants). Similarly, the 6 long-streak sequences were paired with the same target items to give a set of 12 "long-streak context" sequences. In both the short-streak and long-streak contexts, the target item in each context-target pair was always shown second[4]. To ensure generality, three such sets of short-streak context and long-streak context sequences were produced and pseudorandomly assigned to participants. The order of the context-target pairs were randomized for each participant. Examples of the sequences are shown in Table 1.

[4] This paired structure was used to ensure that the contextual items were evenly distributed throughout the set and that participants had always seen at least one contextual item before judging their first target sequence.

The Basketball version of the task was identical, except that the “Heads” and “Tails” outcomes were replaced with “Miss” (M) and “Score” (S), respectively, and the wording of the instructions and response scale were modified accordingly. Participants were pseudorandomly assigned to the Coin and Basketball versions of the task. In total, 218 participants completed the Coin version (110 in the Long context condition and 118 in the Short context condition) and 215 completed the basketball version (107 in the Long context condition and 108 in the Short context condition).

Results and Discussion

The data were coded such that larger numbers indicate greater belief that the current streak will continue. For each participant, the mean judgment was calculated for both context and target sequences. The results are shown in Figure 2.

Consider first the effect of streak length. Between-participant comparisons showed that streaks of 5 outcomes were judged less likely to continue than streaks of one outcome for both Coin and Basketball versions of the task [$t(206.7) = 9.61$, $d = 1.30$, $p < .001$ and $t(166.0) = 8.25$, $d = 1.13$, $p < .001$, respectively]. Similarly, paired-samples t-tests showed that the medium-length streaks of the target sequences were judged more likely to continue than the five-outcome streaks of the long-streak context items (for Coin task: $t(109) = 7.13$, $d = 0.68$, $p < .001$; for Basketball task: $t(106) = 3.06$, $d = 0.33$, $p = .003$) but less likely to continue than the one-outcome streaks at the end of the short-streak context sequences (for Coin task: $t(107) = 8.99$, $d = 1.61$, $p < .001$; for Basketball task: $t(107) = 11.89$, $d = 1.51$, $p < .001$). In short, the data replicate the common finding that longer streaks are judged more likely to end than shorter ones.

More important is the effect of context on the target sequences. Independent samples *t*-tests showed that, for both the Coin and Basketball versions of the task, predictions for the target sequences depended on the other sequences seen during the experiment. Specifically, the three-outcome streaks were judged more likely to continue in the long-streak context than in the short-streak context (for Coin task: $t(198.4) = 5.80$, $d = 0.79$, $p < .001$; for Basketball task, $t(213) = 4.78$, $d = 0.65$, $p < .001$). (The context effect was the same for both the coin flipping and basketball scenarios: A 2 (scenario) x 2 (context) ANOVA on the judgments of the target sequences revealed no main effect of scenario, $F(1,429) = 1.96$, $p = .162$, $\eta_p^2 = .01$ and no context x scenario interaction, $F(1,429) = .16$, $p = .686$, $\eta_p^2 = .00$.)

To check the robustness of these findings, I conducted a supplementary experiment using the basketball task[5]. For participants in the short-context condition, sequences ended with streaks of 1, 2, or 3 (each person judged 6 sequences of each type; 3 ending with a streak of “scores” and three ending with a streak of “misses”); the long-context condition was similar but used streaks of 3, 4, or 5. The order of sequences was randomized for each participant, who indicated their confidence in their predictions on a 6-point scale. (Points 1-3 were labelled Very/Moderately/Slightly confident the next shot will miss; points 4-6 were labelled Slightly/Moderately/Very confident the next shot will score; responses were coded such that larger numbers indicated greater belief in streak continuation). The results mirrored the main experiment: Target streaks of length 3 were judged less likely to continue in the short-context condition (N = 188; M = 2.64, SD = 0.82) than in the long-context condition (N = 188; M = 3.21, SD = 0.70), $t(365.9) = 7.23$, $p<.001$.

In summary, when a moderate streak occurs in a context containing several longer streaks it seems to appear relatively short and is thus judged more likely to continue. By contrast, when the same moderate streak occurs in amongst shorter streaks it seems relatively long and people expect it to end. Thus, the contrast effect found for judgments of randomness in Experiment 1 extended to outcome predictions in Experiment 2.

General Discussion

Individually-presented sequences are not judged in isolation. A moderate alternation rate of 0.5 was judged more likely to indicate a random, mechanical process when presented among of sequences with lower alternation rates than when intermixed with high-AR sequences. Similarly, a terminal streak of three was judged more likely to continue when other sequences in the experiment ended with longer streaks than when they ended with shorter ones (with overall outcome proportions and alternation rates matched). Thus, sequence judgments show substantial context effects (at least in some circumstances), and these effects take the form of judgmental contrast rather than assimilation.

This contrast effect provides one explanation for the variability in sequence judgments noted in the Introduction, and urges caution about attempts to attach special significance to particular sequence parameters. The “rule of three” (Carlson and Shu, 2007),

[5] Full details are available from the author.

for instance, is likely to be an over-simplification; the run length that is perceived as indicative of a meaningful streak will depend on the context established by other, recently experienced sequences, and theories of sequence judgment need to incorporate this kind of parameter variation. The gambler's fallacy, for example, may be the result of the representativeness heuristic (Kahneman & Tversky, 1972), but one needs the additional assumption that recently encountered sequences shape people's prototype for what a random sequence looks like, thereby affecting how representative a particular streak seems to be of a random process. Similarly, the GF may reflect people's treatment of random sequences as being akin to sampling without replacement (e.g., Ayton & Fischer, 2004), but with the caveat that people's mental model of the "urn" from which sampling takes place depends upon recently-experienced sequences. The context effects reported here therefore serve to qualify rather than choose between competing accounts of sequence judgment, although the current data do lend credence to the broader theoretical claim that sequence perceptions and predictions depend upon people's actual experiences with particular types of sequence (e.g., Hahn & Warren, 2009; Wilke & Barrett, 2009).

What mechanism produced the contrast effects seen here? The most obvious interpretation is that they reflect a relative judgment process in which each sequence was compared against the others in the experiment. There are competing ideas about how such relative judgment may proceed. Helson's (1964) adaptation level theory of psychophysical judgment posits the mean of the ensemble as the reference point, whereas Parducci's (1965) range-frequency theory of category judgment assumes that judgments are a based on a compromise of rank position and position in the stimulus range. More recently, Stewart, Chater, and Brown's (2006) Decision by Sampling account takes rank position to be of primary importance to a number of economically important judgments, and manipulating the rank position of target quantities has been found to affect people's preferences in decisions under risk involving both monetary outcomes (e.g., Stewart, 2009) and human lives (Olivola & Sagara, 2009). Finally, in a preliminary study, Altmann and Burns (2005) explored a model in which the current streak was compared against sequences stored in memory using the ACT-R modelling framework (Anderson & Lebiere, 1998); if the current streak is shorter than the one retrieved, a repeat of the current outcome is predicted – giving the contrast effect seen in Experiment 2. It should be possible to choose between these competing possibilities by manipulating the composition of the judgment set (for example, by independently varying a given item's rank position, distance from the mean, and position in the stimulus range), to see which factors influence judgment). In any case, the pronounced context effects found here suggest that models of perceived randomness and outcome prediction can be integrated with broader theories of psychophysical judgment, economic decisions, and memory (see also Olivola & Oppenheimer, 2008).

One question to arise from the current study is: What constitutes the comparison set against which the current sequence is compared? Similarity is likely to be a key determinant, and it may be that the context is overwhelmingly comprised of the other items presented in the experimental session, rather than incidental or superficially different stimuli encountered in the past (e.g., Matthews, 2012). However, there is evidence that context effects can extend beyond these kinds of immediate comparisons. For example, Tresselt (1948) found that the weight judged "medium" differed between weightlifters (who routinely lift heavy objects) and watchmakers (who routinely lift light objects; see also Olivola and Sagara, 2009, for a similar effect of background context on judgment). This raises the intriguing possibility that part of the individual variation in randomness judgments and the tendency to show the gambler's and hot hand fallacies after encountering a given sequence (e.g., Sundali & Croson, 2006) emerges from pre-experimental experience with different types of sequence. It is also worth noting that, even if the context effects are limited to intra-experimental comparisons, there are potentially important implications when, for example, researchers are deciding between within- and between-subject designs.

A related issue concerns whether the contextual effects require overt judgment of the context stimuli. We can ask whether context affects people's *perceptions* of a given sequence, or their use of the response scale. The latter possibility might suggest no effect of contextual stimuli that are not overtly judged – although there is no way of establishing whether participants have covertly labelled the stimuli, and the behavioural outcome (the overt response to the target item) will typically be the same irrespective of the locus of the context effect. More generally, it will be important to investigate alternative sequence presentation formats. Barron and Leider (2010) found that the gambler's fallacy was attenuated when, as in the current experiments, the outcomes forming a sequence were presented simultaneously rather than one at a time. Moreover, they found no GF at all when they replicated Burns and Corpus' (2004) experiment using visually-presented outcomes rather than the original authors' verbal description of the sequence structure. Whether the context effects established in the current work extent to other forms of sequence presentation and judgment task will be a useful line of future research.

Acknowledgements

This research was partially supported by the British Academy. I am grateful to Miu Yee Hung and Robert Baines for their help with data collection during the early stages of this work.

References

Altmann, E.M., & Burns, B.D. (2005). Streak biases in decision making: data and a memory model. *Cognitive Systems Research*, *6*, 5-16.

Anderson, J.R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, New Jersey: Erlbaum.

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*, 1369-1378.

Barron, G., & Leider, S. (2010). The role of experience in the gambler's fallacy. *Journal of Behavioral Decision Making*, *23*, 117-129.

Boynton, D.M. (2003). Superstitious responding and frequency matching in the positive bias and gambler's fallacy effects. *Organizational Behaviour and Human Decision Processes*, *91*, 119-127.

Burns, B.D., & Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand". *Psychonomic Bulletin & Review*, *11*, 179-184.

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

Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from Casinos. *Journal of Risk and Uncertainty*, *30*, 195-209.

Damisch, L., Mussweiler, T., & Plessner, H. (2006). Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied*, *12*, 166-178.

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

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

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

Kwan, V.S.Y., Wojcik, S.P., Miron-shatz, T., Votruba, A.M., & Olivola, C.Y. (2012). Effects of symptom presentation order on perceived disease risk. *Psychological Science*, *23*, 381-385.

Matthews, W.J. (2010). The gambler's fallacy in retrospect: A supplementary comment on Oppenheimer and Monin (2009). *Judgment and Decision Making*, *5*, 133-137.

Matthews, W.J. (in press). How much do incidental values affect the judgment of time? *Psychological Science*.

Matthews, W.J., & Stewart, N. (2009). Psychophysics and the judgment of price: Judging complex objects on a non-physical dimension elicits sequential effects like those in perceptual tasks. *Judgment and Decision Making, 4*, 64-81.

Olivola, C.Y., & Oppenheimer, D.M. (2008). Randomness in retrospect: Exploring the interactions between memory and randomness cognition. *Psychonomic Bulletin & Review¸15*, 991-996.

Olivola, C.Y., & Sagara, N. (2009). Distributions of observed death tolls govern sensitivity to human fatalities. *Proceedings of the National Academy of Sciences, 106*, 22151-22156.

Oskarsson, A.T., Van Boven, L., McClelland, G.H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, *135*, 262-285.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72*, 407-418.

Smith, G. (2003). Horseshoe pitchers' hot hands. *Psychonomic Bulletin & Review*, *10*, 753-758.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology*, *62*, 1041-1062.

Stewart, N., Chater, N., Stott, H.P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General, 132*, 23-46.

Stewart, N., Chater, N., & Brown, G.D.A. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1-26.

Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, *1*, 1-12.

Tresselt, M.E. (1948). The effect of the experiences of contrasted groups upon the formation of a new scale of judgment. *Journal of Social Psychology*, *27*, 209-216.

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

Table 1. Example sequences from Experiments 1 and 2.

Experiment	Sequence Type	Examples
1	Low-AR context	@ @ @ @ @ @ @ @ @ @ # # # # # # # # @ # #
1	Low-AR context	# @ @ @ @ @ @ @ @ # # # # # # # # # @ @ @
1	Target	# @ @ @ # @ @ # @ @ @ # # @ @ # # # # @ #
1	Target	@ @ # # # # @ @ # @ # @ @ # @ @ @ # # # @
1	High-AR context	# @ # @ @ @ # @ # @ # @ # @ # @ # @ # # @
1	High-AR context	@ # @ # @ # @ # @ # @ # # @ # @ @ # @ @ #
2	Short-streak context	T H T H T T T H H T H H H H H T T T H T
2	Short-streak context	H T H H H H T T T H H T H T T T H T T H
2	Target	T T H H T T H H H T H T H H T H H T T T
2	Target	T T H H T T H H H T H T H H T H H T T T
2	Long-streak context	H T T T T H T T T H T H T H T H H H H H
2	Long-streak context	T H H H T H T H H T H H H T H T T T T T

Note: The examples shown for Experiment 2 are from the Coin task. For the Basketball task, T and H were replaced by S and M, respectively.

Figure 1.

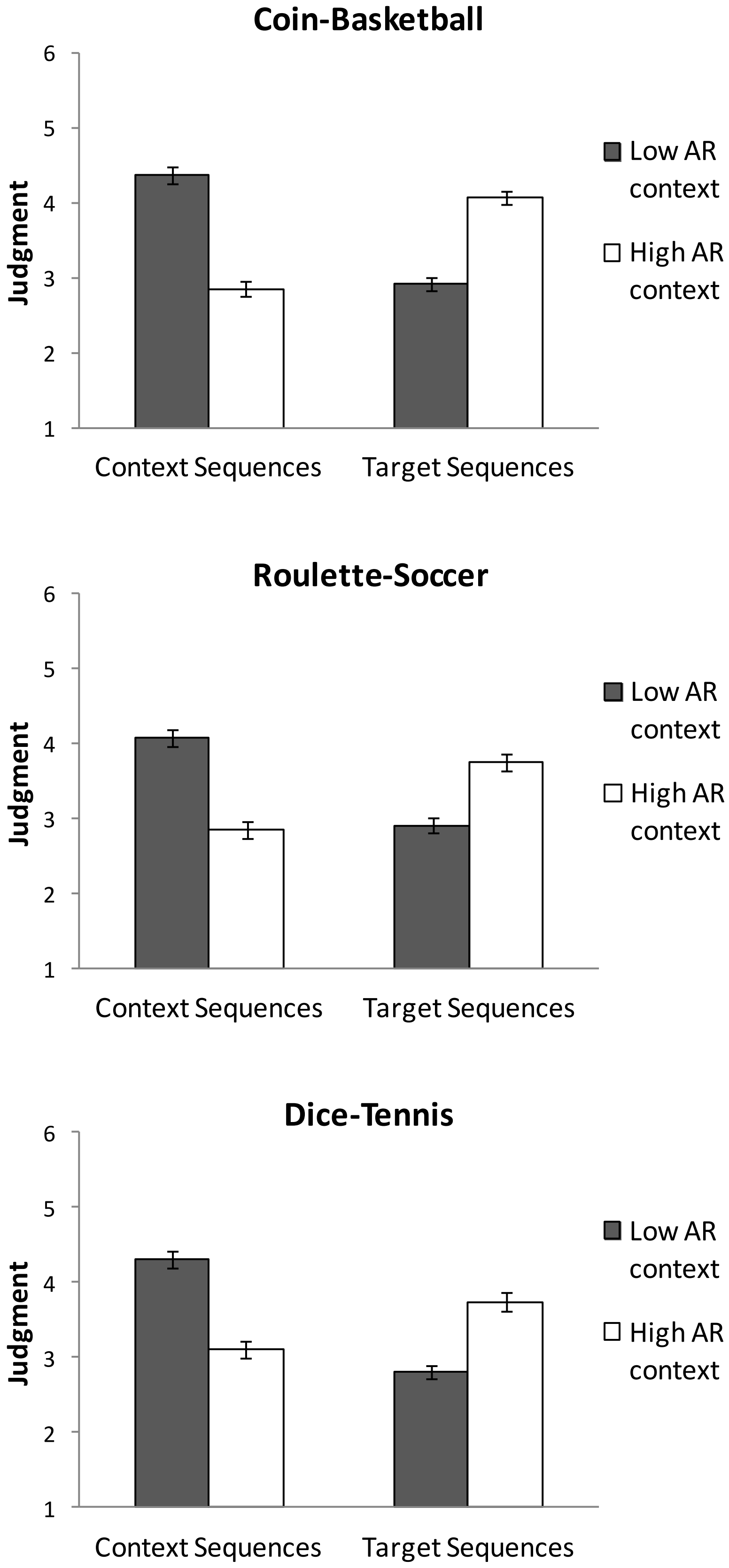
Coin-Basketball
Judgment
6
5
4
3
2
1
Context Sequences
Target Sequences
Low AR context
High AR context
Roulette-Soccer
Judgment
6
5
4
3
2
1
Context Sequences
Target Sequences
Low AR context
High AR context
Dice-Tennis
Judgment
6
5
4
3
2
1
Context Sequences
Target Sequences
Low AR context
High AR context

Figure 1. Results of Experiment 1. The graphs show the mean judgment for each type of sequence in each scenario; larger values indicate greater belief that the sequence was produced by skilled human action (a basketball player, soccer player, or tennis player) and smaller numbers indicate belief that the sequence was produced by a “random” mechanical device (coin flips, roulette wheel spins, or rolling dice). The Target sequences always have an alternation rate of 0.5. In the Low AR context, the context sequences have an AR of 0.15; in the high AR context, the context sequences have an AR of 0.85. Errors bars indicate plus/minus one standard error.

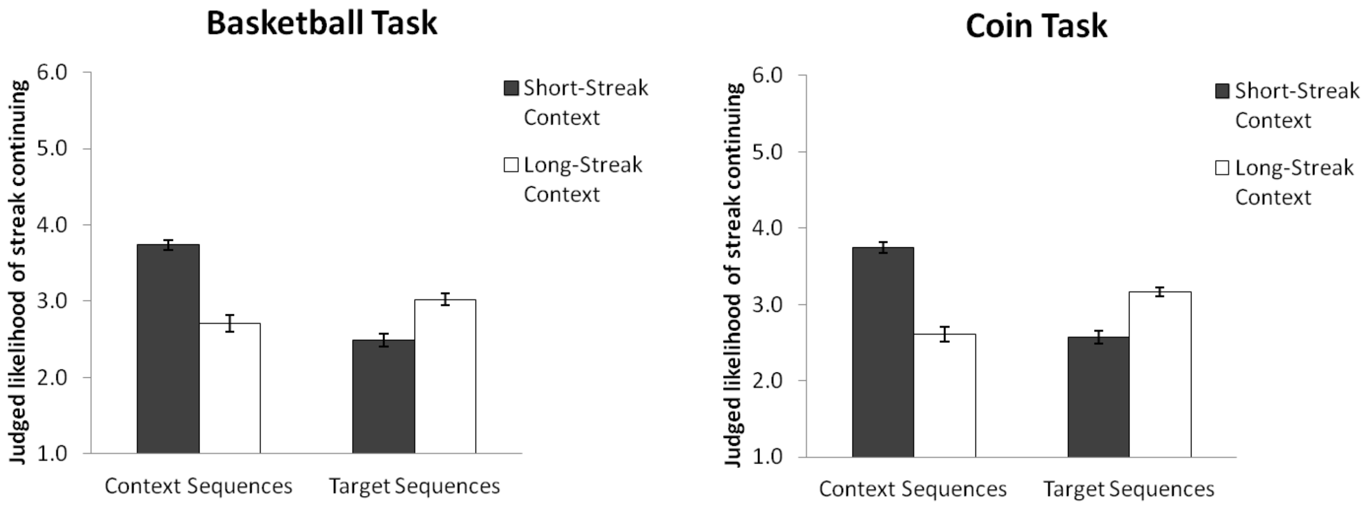

Figure 2. Results of Experiment 2. The graphs show the judged likelihood that the streak at the end of the sequence would continue; larger numbers indicate greater belief in streak continuation. The left and right panels show the results when the sequence was framed as a series of basketball shots and a series of coin flips, respectively. Target sequences always ended in a streak of three. In the short-streak context condition, context sequences ended with a run of one; in the long-streak context condition they ended with a run of five. Error bars show plus/minus one standard error.