

# Randomness in retrospect: Exploring the interactions between memory and randomness cognition

CHRISTOPHER Y. OLIVOLA AND DANIEL M. OPPENHEIMER

*Princeton University, Princeton, New Jersey*

People tend to believe that sequences of random events produce fewer and shorter streaks than is actually the case. Although this error has been demonstrated repeatedly and in many forms, nearly all studies of *randomness cognition* have focused on how people think about random events occurring in the present or future. This article examines how our biased beliefs about randomness interact with properties of memory to influence our judgments about and memory for past random events. We explore this interaction by examining how beliefs about randomness affect our memory for random events and how certain properties of memory alter our tendency to categorize events as random. Across three experiments, we demonstrate an interaction between randomness cognition and three well-established but distinct properties of memory: (1) the reconstructive nature of memory, (2) primacy and recency effects, and (3) duration neglect. Theoretical and practical implications are discussed.

Human reasoning about randomness is systematically biased (Nickerson, 2004). People seem to believe that small samples of random events are representative of the true distribution of possible outcomes (Kahneman & Tversky, 1972).<sup>1</sup> This leads many to fall prey to the classic *gambler's fallacy*: the belief that streaks (i.e., repetitions of an event) are more likely to end than chance would dictate. This erroneous conception of randomness is so strong that when a sequence of events seems to violate it, people often reject the hypothesis that the underlying process is random (Falk & Konold, 1997).

We use the term *randomness cognition* to refer broadly to the set of beliefs, mental representations, schemata, and reasoning processes that people use to think about random events and processes. Research has established that our erroneous beliefs affect the way we generate (Wagenaar, 1972), make predictions about (Croson & Sundali, 2005), and identify (Ayton & Fischer, 2004) random sequences of events; however, almost all studies of randomness cognition have focused on how people think about random events occurring in the present or future. Very little research has examined how people think about past random events.

Studies exploring the connections between memory and randomness cognition have examined how working memory governs the prediction or generation of future random events (see, e.g., Baddeley, Emslie, Kolodny, & Duncan, 1998; Kareev, 1992), assessed how the difficulty of encoding a sequence into memory affects its perceived randomness (Falk & Konold, 1997), or tested memory for events that incidentally involved randomness cognition (recalling choices in competitive games in which randomizing was optimal; Budescu & Rapoport, 1994). To our

knowledge, however, no studies have systematically explored randomness in retrospect—how beliefs about randomness interact with properties of memory to influence the way we judge and remember *past* random events.

This article aims to fill this empirical void. We examine both directions of this relationship—how beliefs about randomness affect what we remember, and how properties of memory affect what we believe is random. We focus on three well-documented properties of memory: its reconstructive nature (Experiment 1), primacy and recency effects (Experiment 2), and duration neglect (Experiment 3). In the experiments we report, participants performed one trial of a serial memory task involving a long sequence of binary events, before recalling the sequence (Experiment 1) or judging its randomness (Experiments 2 and 3).

## EXPERIMENT 1 Reconstructive Memory

Memory is fallible, involving active reconstruction of past events through the use of knowledge and beliefs about the world (Bartlett, 1932). Our memory system stores the essential elements and later fills in the missing bits and pieces via top-down processing, much as our perceptual system functions (Roediger, 1996). Although this is generally efficient, it can cause errors when our beliefs conflict with past realities (Bower, Black, & Turner, 1979). The process of remembering is thus analogous to the formation of judgments, and it may be subject to the same heuristics and biases. Therefore, memory for random sequences might be biased by false beliefs about

randomness. If so, people will make systematic errors in recalling random sequences that resemble errors made when generating random sequences. Namely, when recalling a long, random sequence of events, people will remember the streaks in the sequence as being shorter or less frequent than they actually were. One would expect this effect to be moderated by beliefs about the nature of the sequence, and for those who believe that the sequence is random to exhibit the bias more strongly than those who do not. We tested this hypothesis in Experiment 1.

In Experiment 1, we investigated whether misconceptions about randomness bias memory for random sequences. Participants were led to believe either that a sequence was randomly generated or that it was the product of a complex algorithm with embedded order. Participants who believe that a sequence is randomly generated should retrospectively underestimate the length of streaks, which should negatively impact their recall accuracy, especially for “streaky” sequences. Participants who believe that the sequence is not randomly generated will not be surprised by streaks, and they should not show such a bias.

**Method**

**Participants.** Eighty Princeton undergraduates participated for course credit.

**Procedure.** Participants were asked to memorize a sequence of 25 binary events composed of Xs and Os appearing one at a time on a computer screen (see Figure 1). Each event was presented for 1.5 sec, followed by a half-second delay screen consisting of a multicolored pattern. Each participant was randomly assigned one of two conditions, in a between-subjects design. Participants were not allowed to physically record the sequence as it occurred.

Forty binary random sequences were generated. Two participants—1 per experimental condition—were randomly assigned to each sequence, so that sequence characteristics were identical across conditions.

Participants in the *random-belief* condition were told truthfully that the sequence was simulating the outcomes of fair coin flips and thus consisted of random binary events occurring independently and with equal likelihood. Participants in the *complex-pattern-belief* condition were told that the sequence was generated by an extremely complex algorithm possessing some embedded order. To dissuade participants from searching for a pattern within the sequence as it

unfolded, we showed them a long, fictional mathematical formula and told them that the outcomes were generated by this equation. In reality, both groups were presented with the same set of sequences, all of which were randomly generated.

To make the belief manipulation more convincing, the computer appeared to load a simulator program before displaying the instructions. In the random-belief condition, the screen displayed a drawing of a hand flipping a coin and the words “Coinflip Simulator.” In the complex-pattern-belief condition, the logo was a set of pictures symbolizing complexity (a nautilus shell, a fractal image, a set of connected nodes, and a snowflake) accompanied by the title “Complexity Simulator.” Below each logo, a rotating timer icon appeared for 5 sec to give the illusion that the computer was actually loading the “simulator.”

Participants were told that they would be asked to recall the sequence, in order. Participants first practiced the task with a short sequence of 5 events (OXXOX for all participants) in order to familiarize themselves with the stimuli and procedure for responding. Participants were then shown the 25 events, and they reported the entire sequence to the best of their ability immediately after.

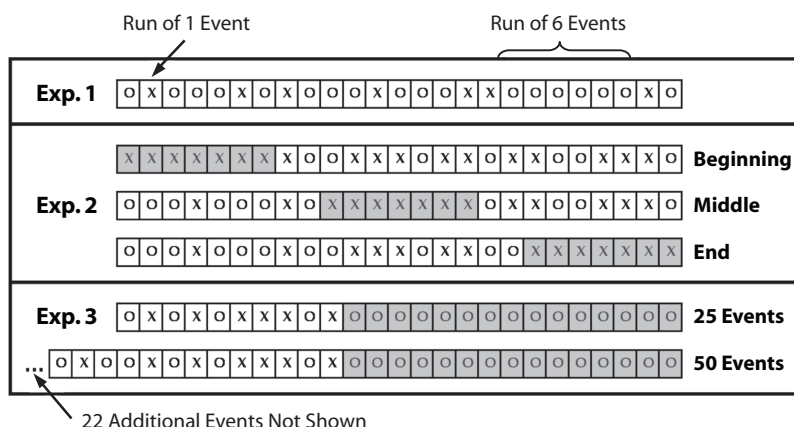
**Results and Discussion**

Paired-samples *t* tests (comparing participants assigned to the same sequence) revealed that participants in the complex-pattern-belief condition recalled streaks as being longer than did those in the random-belief condition [ $t(39) = 2.14, p < .04$ , and  $t(39) = 2.48, p < .02$ , for recalled mean<sup>2</sup> and maximum streak lengths, respectively]. To test for systematic biases in recall, we focused on two types of deviations: (1) the difference between the mean lengths of actual and recalled streaks ( $\Delta_{avg}$ ) and (2) the difference in length between the longest actual streak and the longest recalled streak in a sequence ( $\Delta_{max}$ ). Thus, positive numbers signified overestimation of mean or maximum streak length, whereas negative values indicated underestimation:

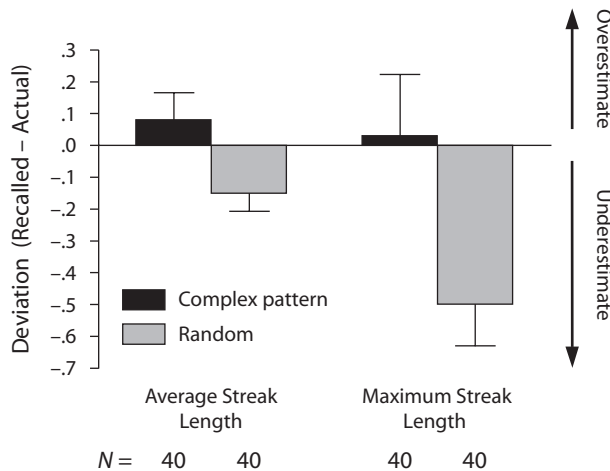
$$\Delta_{avg} = \text{avg}(\text{recalled}) - \text{avg}(\text{actual});$$

$$\Delta_{max} = \text{max}(\text{recalled}) - \text{max}(\text{actual}).$$

The results are presented in Figure 2. The deviation between recalled and actual mean streak lengths ( $\Delta_{avg}$ ; see note 2) differed significantly across belief conditions [ $t(78) =$



**Figure 1.** Examples of the sequences used in each experiment. Inserted streaks have been highlighted for the reader.



**Figure 2.** Average differences between recalled and actual streak lengths as a function of belief condition (Experiment 1). Error bars represent standard error.

2.18,  $p < .04$ ,  $d = .49$ ). Participants in the random-belief condition retrospectively underestimated the length of streaks [ $M = -.15$ ,  $SD = .36$ ;  $t(39) = 2.66$ ,  $p < .02$ ]. In contrast, no significant bias was found for participants in the complex-pattern-belief condition [ $M = .07$ ,  $SD = .55$ ;  $t(39) = 0.86$ ,  $p > .39$ ]. The deviation between recalled and actual maximum streak lengths ( $\Delta_{\max}$ ) also differed across belief conditions [ $t(78) = 2.25$ ,  $p < .03$ ,  $d = .51$ ]. Participants in the random-belief condition retrospectively underestimated the length of the longest streak [ $M = -.50$ ,  $SD = .82$ ;  $t(39) = 3.87$ ,  $p < .0004$ ]. Participants in the complex-pattern-belief condition, however, did not show this bias [ $M = .03$ ,  $SD = 1.23$ ;  $t(39) = 0.13$ ,  $p > .89$ ].

We also looked at recall accuracy, measured as the proportion of events that were correctly recalled and correctly located in the sequence. Although mean accuracy was slightly higher in the random-belief condition (66%) than in the complex-pattern-belief condition (63%), this difference was not significant [ $t(78) = 0.88$ ,  $p > .38$ ]. For participants in the complex-pattern-belief condition, there was a significant positive correlation between actual mean streak length<sup>3</sup> and accuracy [ $r(40) = .49$ ,  $p < .002$ ] and between actual maximum streak length and accuracy [ $r(40) = .46$ ,  $p < .004$ ]. Regressing accuracy on actual mean streak length revealed that each unit increase in average streak length was associated with a 14% boost in accuracy. For participants in the random-belief condition, however, accuracy did not correlate with mean streak length [ $r(40) = -.05$ , n.s.], nor did it correlate with maximum streak length [ $r(40) = -.11$ , n.s.]. Regressing accuracy on average (or maximum) streak length, belief condition (a dummy variable), and their interaction revealed that only the interaction between streak length and belief condition was significant [ $t(76) = 2.42$ ,  $p < .02$ ,  $\beta = .37$ , and  $t(76) = 2.57$ ,  $p < .02$ ,  $\beta = .39$  for average and maximum streak length, respectively; both main effect  $t$  values  $< 1$ ], suggesting that the effect of streak length on accuracy was fully moderated by the belief condition.

We were concerned that participants might not have remembered the sequences at all and had resorted to guessing. Such generation of novel sequences would reflect their beliefs about randomness without being a memory bias; however, complex-pattern-belief respondents showed a strong correlation between actual streak length and recall accuracy. This is contrary to the near-zero correlation we would have expected if participants were generating novel sequences. Furthermore, guessing should have produced chance-level recall accuracy ( $\sim 50\%$ ), but our participants showed above-chance recall, even without the benefits of primacy and recency: Recall accuracy for the middle 15 events (ignoring the first 5 and last 5 events) significantly exceeded chance in the random-belief [62%;  $t(39) = 4.81$ ,  $p < .00003$ ] and complex-pattern-belief [58%;  $t(39) = 3.01$ ,  $p < .005$ ] conditions.

Participants who believed that the sequence that they had seen was randomly generated remembered its streaks as being shorter than they actually were, whereas participants who believed the sequence to be nonrandom showed no such bias in recall. Furthermore, although accuracy increased with streak length for participants in the complex-pattern-belief condition, no such correlation was found for participants in the random-belief condition. This suggests that participants' beliefs about randomness led them to underestimate the length of streaks, thereby countering the recall benefits provided by "chunky" sequences containing long streaks. This experiment demonstrates how one's belief that a sequence is random can bias memory. In the subsequent two experiments, we examined the inverse: how properties of memory can bias our belief that a sequence is random.

## EXPERIMENT 2 Primacy and Recency

The serial recall literature has established that the first few and last few events in a temporal sequence are usually the easiest to remember—the primacy and recency effects (Murdoch, 1962). Consider how these effects might impact beliefs about whether a past sequence is random. Sequences that contain long streaks are judged to be less random than sequences with shorter streaks (Kahneman & Tversky, 1972). When making this judgment from memory, however, people may be less likely to categorize a sequence as random if a streak occurs at the beginning or end of the sequence, where it is likely to be recalled, than if it happens in the middle of the sequence, where it is more likely to be forgotten. We tested this hypothesis in Experiment 2.

To test whether judgments of sequence randomness would be influenced by primacy or recency effects, we asked participants to memorize a pseudorandom sequence with a long streak inserted at the beginning, middle, or end. They then judged whether the sequence was random or not. One would expect participants shown a sequence with a streak in the middle to be more likely to retrospectively categorize that sequence as random than participants shown a sequence with a streak at the beginning or the end. In contrast, one would expect that participants

who do not have to rely on memory to make their judgment would not display such a bias.

**Method**

**Participants.** Four hundred eighty-six Princeton undergraduates were recruited.

**Stimuli and Design.** Twenty pseudorandom binary sequences of 25 events were generated (see Figure 1).<sup>4</sup> For each sequence, a mirror sequence was also created (e.g., the mirror sequence for “XOOXO . . .” is “OXXOX . . .”). The first, middle, or last 5 or 7 events were replaced with a streak of five or seven Xs or Os, producing a total of 240 sequences (20 sequences × 2 mirrors × 3 streak locations × 2 streak lengths). When necessary, the events directly preceding or following the streaks were altered in order to make streaks exactly 5 or 7 events long. Participants were randomly assigned to one of these 240 sequences.<sup>5</sup>

Participants in the memory condition saw the events one at a time on a computer screen and attempted to memorize the sequence, knowing only that they would later be asked something about it (but not exactly what they would be asked). Each event was presented for 1 sec, followed by a half-second delay screen. Participants in the nonmemory condition saw the entire sequence all at once on paper.

Immediately following the sequence, participants were asked to judge whether it was randomly generated. Those in the memory condition had to make their judgments from memory, whereas those in the nonmemory condition could consult the sequence while making their judgments. Identical sequence stimuli were used in both groups.

**Results and Discussion**

Results were similar for sequences with streaks of five or seven events, so we collapsed across streak length for all subsequent analyses. We examined the proportion of participants who judged a sequence (retrospectively or directly) to be random as a function of streak position (see Figure 3).

Among participants in the memory condition, randomness judgments differed significantly between streak locations [ $\chi^2(2) = 7.08, p < .03, \phi = .17$ ]. Sequences were

more frequently categorized as random when the streak was in the middle than when it was in the beginning [ $\chi^2(1) = 6.02, p < .02, \phi = .19$ ] or the end [ $\chi^2(1) = 5.03, p < .03, \phi = .18$ ]; however, randomness judgments did not differ when the streak was in the beginning versus when it was at the end of the sequence [ $\chi^2(1) = 0.05, p > .82$ ].

In contrast, participants in the nonmemory condition did not differ in their randomness judgments between streak locations [ $\chi^2(2) = 0.41, p > .81$ ]. This suggests that the effects of streak location were mainly due to properties of memory rather than to the perception of sequences per se.

Participants’ judgments of sequence randomness were influenced by primacy and recency effects; those shown a sequence with a streak in the middle were more likely to retrospectively categorize the sequence as random than were participants who saw the same streak occur at the beginning or end. When they did not have to rely on memory, however, these differences disappeared. This suggests that the serial position of streaks influences randomness judgments from memory. In the final experiment, we examined whether sequence length would also affect randomness judgments.

**EXPERIMENT 3  
Duration Neglect**

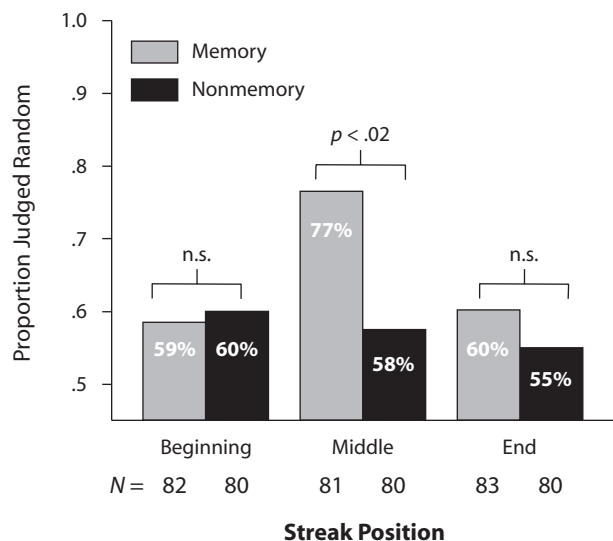
Although people can remember key moments in an extended episode, they are often insensitive to its duration (Fredrickson & Kahneman, 1993). This *duration neglect* has implications for whether we judge a sequence of events to be random. Research suggests that when streaks occur in long sequences, the sequences are considered more random than when the same streaks occur in shorter sequences (Oppenheimer & Monin, 2008); however, this difference should disappear when randomness judgments are made from memory. Recall of extended temporal sequences should be unaffected by the length of those sequences. Thus, when judging randomness from memory, people should fail to correct for sequence length. We tested this hypothesis in Experiment 3.

If duration neglect applies to retrospective randomness judgments, then we expect participants who are asked to memorize long sequences and participants who are asked to memorize short sequences to be equally likely to retrospectively categorize the sequences as random. In contrast, we expect participants judging the randomness of sequences they can see directly to be sensitive to length. That is, longer sequences will attenuate their belief that the presence of a streak indicates nonrandomness.

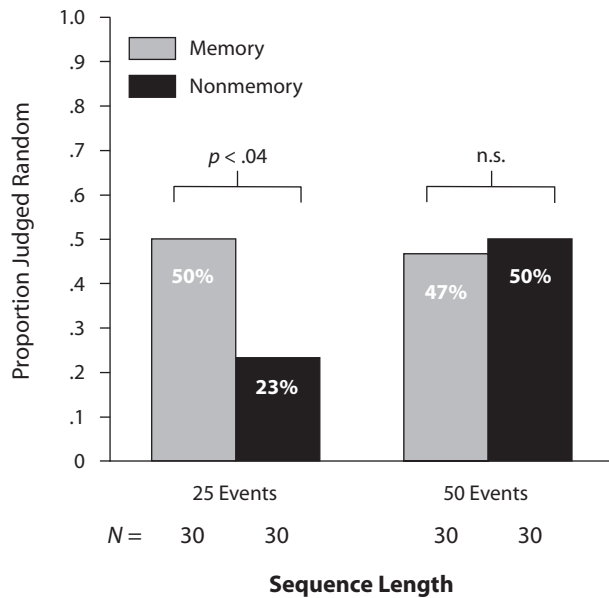
**Method**

**Participants.** One hundred twenty Princeton undergraduates were recruited.

**Stimuli and Design.** Thirty pseudorandom binary sequences were generated (see Figure 1, and see note 4). Half were 25 events long (short sequences), and half were 50 events long (long sequences). Mirror sequences were also created. The last 15 events in each sequence were replaced with a streak of Xs or Os. When necessary, events directly preceding the long streaks were altered in order to keep these streaks exactly 15 events long. Participants were



**Figure 3.** Proportion of sequences judged to be random as a function of streak position and memory condition (Experiment 2). The *p* values above the bars comparing memory conditions are based on chi-square tests.



**Figure 4.** Proportion of sequences judged to be random, as a function of sequence length and memory condition (Experiment 3). The  $p$  values above the bars comparing memory conditions are based on paired-samples Wilcoxon signed-ranks tests (comparing participants in each memory condition assigned to the same sequence).

randomly assigned to one of these 60 sequences. Identical sequence stimuli were used in the memory and nonmemory conditions. The procedures were identical to those in Experiment 2.

### Results and Discussion

We examined the proportion of participants who judged a sequence (retrospectively or directly) to be random as a function of sequence length (see Figure 4).

Participants in the memory condition were as likely to retrospectively judge a short sequence to be random as they were to retrospectively judge a long sequence to be random [ $\chi^2(1) = 0.07, p > .79$ ]. In contrast, participants in the nonmemory condition were more likely to judge sequences with 50 events to be random than to judge sequences with 25 events to be random [ $\chi^2(1) = 4.59, p < .04, \phi = .28$ ].

Retrospective judgments of sequence randomness were not influenced by sequence length; long and short sequences containing a streak were equally likely to be judged random. In contrast, sequence length did influence randomness judgments when these judgments were not memory based, which meant that longer sequences were more likely to be judged random. This suggests that duration neglect in memory influences randomness cognition.

### GENERAL DISCUSSION

The results of the present three experiments support the notion that randomness cognition and properties of memory interact. In Experiment 1, participants recalled a random sequence of events. As predicted by the reconstructive property of memory, participants who (correctly) believed that the sequence was randomly generated sys-

tematically underestimated streak lengths, whereas those who believed that the sequence was nonrandomly generated did not show this bias. In Experiment 2, participants relying on memory were less likely to categorize sequences as random when streaks occurred in the beginning or end (where they would be better remembered) than to do so when streaks occurred in the middle. In contrast, no such difference was found for participants who did not rely on memory. The results of these first two experiments are surprising, given the evidence that people are particularly sensitive to the occurrence of streaks (Falk & Konold, 1997). Indeed, one could have predicted a “prominence effect,” in which the salient nature of streaks would make them easy to remember, regardless of their location, and especially when the sequence was thought to be random. This hypothesis was not supported, however. In Experiment 3, participants relying on memory were as likely to categorize long sequences containing a given streak as random as they were to do so for short sequences, failing to adjust for sequence length. In contrast, participants who did not rely on memory were more likely to categorize the longer sequences as random. Although duration neglect has been shown to occur for affective episodes (see, e.g., Fredrickson & Kahneman, 1993), these results are the first to show that it can also occur with nonaffective experiences. In all three experiments, responses were elicited *immediately* after sequence presentation. Thus, we can expect even larger biases as the delay until a response increases.

These data extend the literature on randomness cognition and probabilistic reasoning by presenting a closer look at how people think about past random events. It also adds to a growing literature on the parallels between biases in prediction and biases in memory. For example, people exhibit the same biases when attempting to predict affective states as those they exhibit when attempting to remember affective states (Wirtz, Kruger, Scollon, & Diener, 2003). Our results demonstrate that the same is true of randomness cognition. In Experiment 1, we found that biases previously shown to influence predictions about future random events also impact memory for past random sequences, producing similar errors.

These experiments have interesting implications for how people deal with randomness in the real world. For example, people have difficulty learning to overcome the *gambler's fallacy*, even when real money is at stake and they receive repeated feedback that contradicts their beliefs about random processes (Croson & Sundali, 2005); however, our results show that memory for random events is itself shaped by beliefs about randomness. Thus, gamblers who witness evidence contradicting their intuition may later falsely recall the outcome of the game as being closer to their initial belief. As a result, corrective feedback may be less effective in reducing the occurrence of the gambler's fallacy than one would hope.

These experiments can also inform us about when past events are categorized as random. In finance, for example, investors' beliefs that a particular stock's price is fluctuating randomly might be especially sensitive to sustained increases or decreases in value if these runs occur at the

beginning or end of a trading period. In contrast, investors might be relatively insensitive to the length of time that precedes a run of unidirectional price shifts when deciding whether this run indicates a trend. Results from nonmemory participants suggest that one remedy for these tendencies may be to rely on visual summaries of stock activity.

Although memory and judgment have often been considered distinct from each other, the parallels and interactions between them suggest that their juxtaposition is fertile ground for research. Until we incorporate the role of memory processes more fully, our understanding of randomness cognition remains incomplete.

#### AUTHOR NOTE

The authors thank Peter Ayton, Ruma Falk, David Hardman, Reid Hastie, Mark McDaniel, Benoît Monin, Carey Morewedge, Ray Nickerson, Eldar Shafir, Neil Stewart, and an anonymous reviewer for their helpful suggestions, as well as Naomi Gilens, Isaiah Miller, and Jeff Zemla for their excellent research assistance. Correspondence concerning this article should be addressed to C. Y. Olivola, Department of Psychology, Princeton University, Green Hall, Princeton, NJ 08540 (e-mail: colivola@princeton.edu).

#### REFERENCES

- AYTON, P., & FISCHER, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*, 1369-1378.
- BADDELEY, A. [D.], EMSLIE, H., KOLODNY, J., & DUNCAN, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology*, *51A*, 819-852.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.
- BUDESCU, D. V., & RAPOPORT, A. (1994). Subjective randomization in one- and two-person games. *Journal of Behavioral Decision Making*, *7*, 261-278.
- CROSON, R., & SUNDALI, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk & Uncertainty*, *30*, 195-209.
- FALK, R., & KONOLD, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*, 301-318.
- FREDRICKSON, B. L., & KAHNEMAN, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality & Social Psychology*, *65*, 45-55.
- KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430-454.
- KAREEV, Y. (1992). Not that bad after all: Generation of random sequences. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 1189-1194.
- MURDOCK, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *62*, 482-488.
- NICKERSON, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Mahwah, NJ: Erlbaum.
- OPPENHEIMER, D. M., & MONIN, B. (2008). *The retrospective gambler's fallacy*. Manuscript in preparation.
- ROEDIGER, H. L., III (1996). Memory illusions. *Journal of Memory & Language*, *35*, 76-100.
- WAGENAAR, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, *77*, 65-72.
- WIRTZ, D., KRUGER, J., SCOLLON, C. N., & DIENER, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science*, *14*, 520-524.

#### NOTES

1. See Ayton and Fischer (2004) for an alternative account.
2. Calculated using runs of two events or more.
3. These secondary analyses included runs of one event (see Figure 1), because such runs contribute to sequence complexity and, therefore, to recall accuracy.
4. Pseudorandom sequences were generated using a randomizing function that constrained runs to a maximum of three events.
5. Six additional participants in the memory condition were accidentally assigned to the same sequences as previous participants.

(Manuscript received August 6, 2007;  
revision accepted for publication March 31, 2008.)