The Noncognitive Origins of the Gambler’s Fallacy

Yanlong Sun,[1*] [. . .], Hongbin Wang [2*]

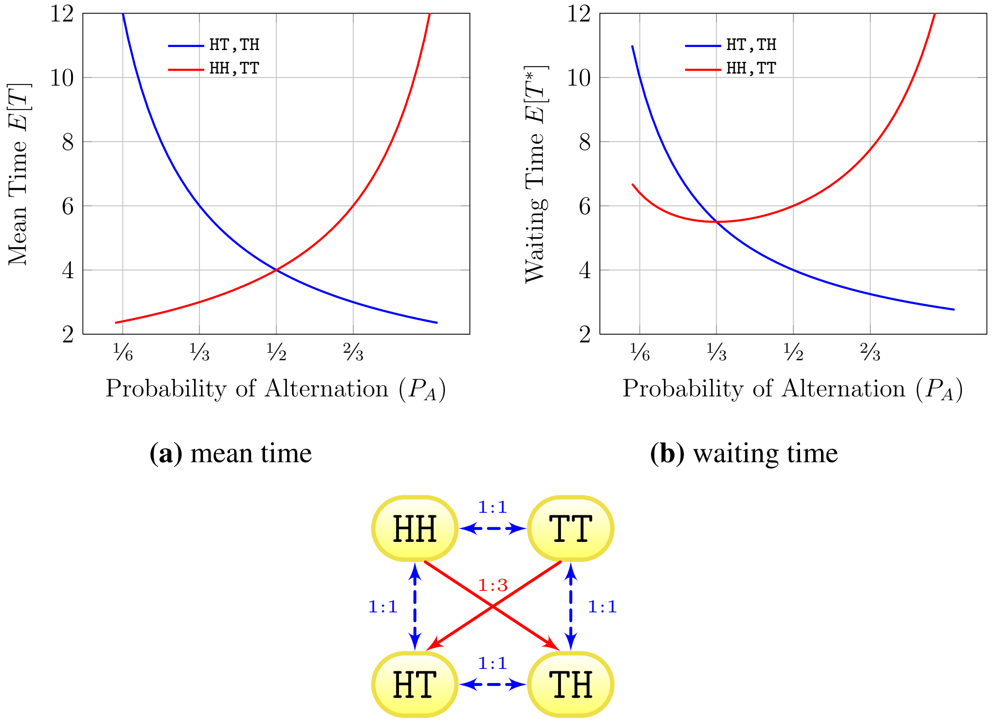
[1,2] University of Texas Health Science Center at Houston,
7000 Fannin Street, Houston, TX 77030, USA

[*]To whom correspondence should be addressed
E-mail: Yanlong.Sun@uth.tmc.edu, Hongbin.Wang@uth.tmc.edu

If the gamblers fallacy has originated from the representativeness bias in a belief-updating structure, what is the origin of the bias and structure in the first place?

The gambler’s fallacy—a belief that chance is a self-correcting process where a deviation in one direction would induce a deviation in the opposite direction—has been a notorious showcase of human irrationality in probabilistic reasoning (*1*). For decades, this fallacy is thought to have originated from a cognitive bias, the “representativeness heuristic”, a belief of the “law of small numbers” that small samples are highly representative of the populations from which they are drawn (*1, 2*). Computationally, representativeness has been defined by a Bayesian belief-updating structure for differentiating hypotheses given different samples of input data (*3–6*). However, it remains controversial how the representativeness bias and the belief-updating structure have originated in the first place (*7–9*).

Here we show that without a predefined hypothesis structure, biases underpinning the gambler’s fallacy can emerge by simply capturing the statistical structures in the learning environment. We used a biologically realistic and minimally configured simple recurrent model that learns to re-encode sequential binary data through unsupervised learning rather than to classify any temporal patterns (*10, 11*). Yet a dissociation of temporal patterns naturally emerged as the consequence of inhibitory competition between representations (*11*) in response to the statistical structures produced by a truly random process (e.g., fair coin tossing) (*12–14*). To our knowledge, this is the first demonstration of a possible brain machinery in accounting for the origin of the gambler's fallacy without resorting to any presumed biases or hypothesis structures. Our findings indicate that cognitive biases in overt behavior can emerge early and locally at the level of perceptual analyses, and, neurons' sensitivity to the statistical structures in the learning environment is the key in bridging the gap between neurons and behavior.



(a) mean time

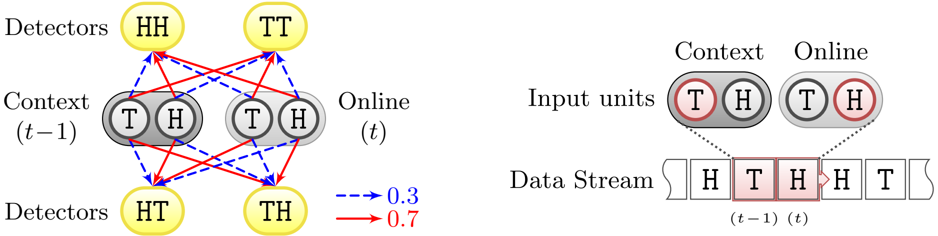
(b) waiting time

(c) precedence ($P_A = 1/2$)

Figure 1 Statistical structures in the learning environment. **(a)** and **(b)**: A pattern's mean time (frequency) and waiting time (delay) as the functions of the probability of alternation P_A (the probability that the current outcome is different from the previous one). **(c)**: The pairwise precedence relationship between patterns when $P_A = 1/2$. In spite of nontransitivity (e.g., HH and HT are equally likely to precede each other), the repetition patterns are on average the mostly delayed (they can at most tie with another pattern). It can be shown that this delayed property holds for all patterns of length greater than or equal to two (*15*).

In sequences generated by a random process, there can be fundamentally different types of statistical structures regarding the occurrences of local samples or substrings (hence referred to as "patterns"). In particular, the *mean time* of a pattern measures how often the pattern occurs in the global sequence, and the *waiting time* measures when a pattern will first occur since the beginning of the observation (*16*). For example, when tossing a unbiased coin independently, "repetition" patterns (HH or TT, i.e., a head following a head or a tail following a tail) and "alternation" patterns (HT or TH) are equally likely to occur in every 4 tosses. However, it takes on average 4 tosses to observe the first occurrence of HT or TH, but 6 tosses for the first occurrence of HH or TT (see **Figure 1**). In addition, the waiting time statistics also manifest as the pairwise precedence between different patterns. For example, the odds are 3 to 1 that one is to first encounter TH than to first encounter HH. On average, the repetition patterns are the most "delayed" patterns.

There has been a growing speculation that people's intuition about random process (i.e., subjective randomness) is biased by the statistical structures in the learning envionment (*12–14, 17, 18*) (also see **Supplementary Material** for a brief review of existing theories). Specifically, the gambler's fallacy has been interpreted as the result of the representativeness bias, in which repetition patterns are *underrepresented* thus less representative of a truly random process than alternation patterns (*1, 2*). And, it was suggested that the bias may have actually originated from the the waiting time rather than mean time statistics (*13*).



(a) model structure

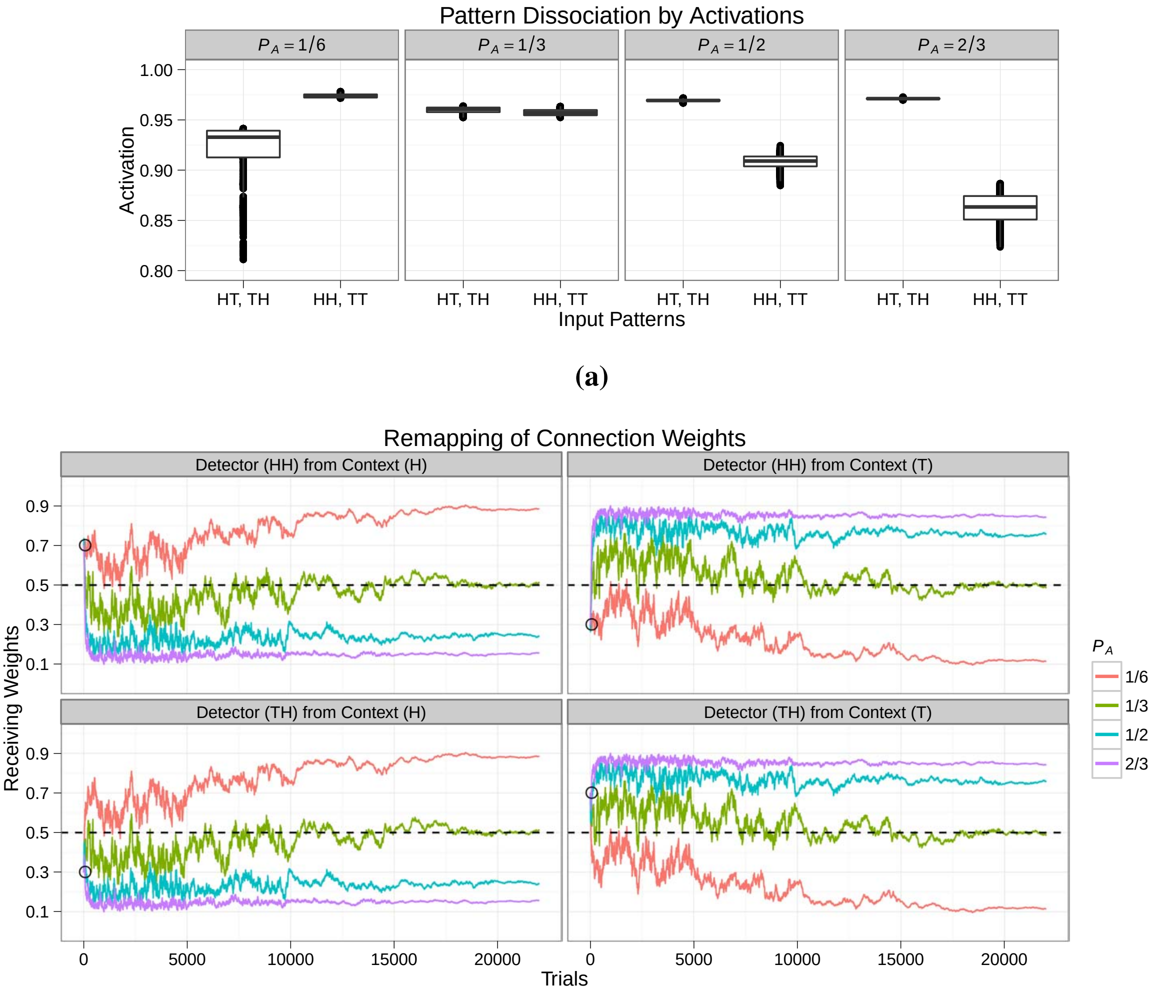
(b) sample input data

Figure 2 A minimal model of temporal pattern dissociation. **(a)** Architecture of the simple recurrent network. The model has two levels of layers, input ("online" and "context") and hidden ("detectors"). At any time, either the H (head) or the T (tail) unit on the "online" layer is active, representing the outcome of tossing a coin. The "context" layer keeps a copy of the activations of the online layer at time $(t-1)$. The hidden layer consists of 4 detectors. The initial receiving weights on the hidden layer are controlled to exhaust all possible combinations of connection biases so that each of the input patterns of length 2 initially has an equally unbiased representation on the hidden layer. **(b)** A sample segment of the input data stream. At each level of the probability of alternation, the model was trained for 300 epochs, each epoch consisted of a newly generated sequence of 100 trials, and each trial was a sequentially independent or dependent coin toss).

Here we set out to find whether a biologically realistic neural network model (*11*) would show sensitivity to such statistical structures. **Figure 2** shows the architecture of the model and training schemes (see **Supplementary Material** for simulation details). At the input level, the online and context layers respectively provided the current (t) and contextual ($t-1$) outcomes from sequences of binary events. At the deeper level, four units on a single hidden layer ("detectors") re-encoded the input data by maximizing the likelihood of reconstructing the input. Crucially, learning concerned only efficient encoding of input and not pattern dissociation, since no teaching signals regarding temporal patterns were provided. The initial receiving weights to the hidden layer were set *symmetrically* so that the model was initially *unbiased* in representing each of the 4 possible patterns of length 2. The controlled weights allowed controlled experiments by exhausting all possible combinations of the connection biases thus eliminating variations produced by random weights (see **Supplementary Material** for the results from different sets of initial weights).

We were interested in how the model would capture the statistical structures shown in **Figure 1**, and consequently, whether consistent biases would arise leading to the representativeness bias and the gambler's fallacy. Since our focus was on the learning environment, we used all default values for the model parameters (*11*) and only manipulated the probability of alternation P_A when generating the input sequences.

Figure 3 shows the main simulation results. We first notice that after being trained with truly random sequences (i.e., $P_A = 1/2$ in independent fair coin tossing), the averaged activations on the hidden layer were significantly lower when the current inputs consisted of repetition patterns (`HH` or `TT`) than that of alternation patterns (`HT` or `TH`) (see **Supplementary Material** for tests of statistical significance). That is, in spite of the initially unbiased representations of all patterns and the equal training frequency (i.e., the mean time is the same for all patterns at $P_A = 1/2$), repetition patterns eventually were significantly underrepresented than alternation patterns. The updating trajectories of the connection weights (**Figure 3b**) revealed that at $P_A = 1/2$, whereas all connection weights from the online layer stayed at the same side of 0.5, the connection weights from the context layer to the `HH` detector had changed to the opposite sides of 0.5, effectively "switching" the `HH` detector into a `TH` detector (but not a `HT` detector). Similarly, the `TT` detector had switched to an `HT` detector (but not a `TH` detector). The directions of these switches corresponded exactly to the pairwise precedence relationship depicted in **Figure 1c**. These observations indicated that the model was sensitive to the delay rather than the frequency of pattern occurrences. This hypothesis was confirmed by simulations at different P_A levels. Particularly, at $P_A = 1/3$, all patterns had the same waiting time but different mean time, so that the model eventually learned to be indifferent to the contextual information (connection weights from the context layer were close to 0.5), resulting in unbiased activations for all patterns. At



(a)

(b)

Figure 3 Pattern dissociation at different levels of the probability of alternation P_A. **(a)**: Dissociation by plus-phase activation values on the hidden layer. Box plots represent distribution quantiles. **(b)**: The remapping of connection weights from the context layer to the `HH` and `TH` detectors on the hidden layer, where a weight either stayed at the same side or switched to the opposite side of 0.5 (the initial values are marked by empty circles). All connection weights from the online layer stayed at the same side of 0.5 and are not shown here. The weight update of the `HH` detector was in the same fashion as the `TT` detector, and `TH` as `HT`.

$P_A = 1/6$, it was the alternation detectors' turn to be underrepresented and switched to the repetition detectors. Moreover, we also tested models with only one particular pattern detector (i.e., a single unit hidden layer). Regardless the initial pattern preference set by the connection weights, the model would eventually react indifferently to any patterns. This result indicated that the inhibitory competition between multiple detectors was required for pattern dissociation. Overall, it appeared that pattern dissociation—namely, different biases for representing either repetition or alternation patterns—emerged from temporal reconstructions of the input data, in which the model "reoriented its attention" to the past information (i.e., updating connection weights from the context layer). And, the driving force behind such attention reorientation was the waiting time statistics in the input data and the competitions between pattern detectors.

In summary, the simulation results showed that the model had learned to capture the waiting time rather than the mean time statistics in the learning environment. Especially when the underlying process was truly random (i.e., independent coin tossing), the model showed significant underrepresentation for repetition patterns, consistent with the representativeness bias underpinning the gambler's fallacy (*1, 2*). For example, **3** shows that at $P_A = 1/2$, the model would be more likely to predict a head following a tail (`TH`) rather than a repetition of two heads (`HH`), despite the sequential independence of events. In terms of representativeness bias, the model had learned that alternation patterns were more representative of a process under $P_A = 1/2$, and repetition patterns were more representative of a process under $P_A = 1/6$. Critically, such bias emerged through unsupervised learning without any pre-defined hypothesis structures (i.e., the model was initially symmetrically structured and was not provided with any priors on how much a pattern would be in favor of one hypothesis over another).

It is noted that our model shown here only addressed purely bottom-up learning mechanisms without implementing any top-down learning or higher-level representations (e.g., beliefs, goals). Nevertheless, it is very plausible that early and locally emergent biases need to be maintained at a higher and global level, and, higher-level representations are responsible for simultaneous analyses through a hierarchical structure of abstractions (*6, 19*). That is, we take the critical issue as "not whether to start at the top or at the bottom, but whether to start at the beginning" (*20*). Given the simplicity of our model, one far-reaching implication is that cognitive biases and structures of abstractions can emerge early and locally at the level of perceptual analyses, and, neurons' sensitivity to the environmental statistics is the key in bridging the gap between neurons and behavior.

References and Notes

1. A. Tversky, D. Kahneman, *Science* **185**, 1124 (1974).

2. T. Gilovich, R. Vallone, A. Tversky, *Cognitive Psychology* **17**, 295 (1985).

3. G. Gigerenzer, U. Hoffrage, *Psychological Review* **102**, 684 (1995).

4. T. L. Griffiths, J. B. Tenenbaum, *Proceedings of the 23rd annual conference of the cognitive science society*, J. D. Moore, K. Stenning, eds. (Lawrence Erlbaum Associates, Mahwah, NJ, 2001), pp. 398–403.

5. J. B. Tenenbaum, T. L. Griffiths, *Proceedings of the 23rd annual conference of the cognitive science society*, J. D. Moore, K. Stenning, eds. (Lawrence Erlbaum Associates, Mahwah, NJ, 2001), pp. 1036–1041.

6. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, *Science* **331**, 1279 (2011).

7. G. Gigerenzer, *Psychological Review* **103**, 592 (1996).

8. J. Lyons, D. J. Weeks, D. Elliott, *Frontiers in Psychology* **4 (Article 72)** (2013).

9. J. L. McClelland, *et al.*, *Trends in Cognitive Sciences* **14**, 348 (2010).

10. J. L. Elman, *Cognitive Science* **14**, 179 (1990).

11. R. C. O'Reilly, Y. Munakata, M. J. Frank, T. E. Hazy, Contributors, *Computational Cognitive Neuroscience* (Wiki Book, 1st Edition, URL: http://ccnbook.colorado.edu, 2012).

12. Y. Sun, R. D. Tweney, H. Wang, *Psychological Review* **117**, 697 (2010).

13. Y. Sun, H. Wang, *Cognitive Psychology* **61**, 333 (2010).

14. Y. Sun, H. Wang, *Judgment and Decision Making* **5**, 124 (2010).

15. R. L. Graham, D. E. Knuth, O. Patashnik, *Concrete mathematics* (Addison-Wesley, Reading MA, 1994).

16. S. M. Ross, *Introduction of probability models* (Academic Press, San Diego, CA, 2007), 9th edn.

17. U. Hahn, P. A. Warren, *Psychological Review* **116**, 454 (2009).

18. D. M. Oppenheimer, B. Monin, *Judgment and Decision Making* **4**, 326 (2009).

19. Y. Munakata, *et al.*, *Trends in Cognitive Sciences* **15**, 453 (2011).

20. G. T. Altmann, *Trends in Cognitive Sciences* **14**, 340 (2010).