A confirmation bias in perceptual decision-making: generalization to abstract cognition

Ruolan Li, Ankani Chattoraj, Ralf Haefner

Confirmation bias, the phenomenon that people are biased towards their previously formed beliefs, is prevalent in our daily experience and the scientific literature. One manifestation of the confirmation bias is a preference for early evidence, shown by more weight given to information earlier in time (i.e., a *primacy effect*). Studies suggest that a primacy effect is present in some settings of perceptual decision-making [1] [2] but not always [3]. Lange et al. (2018) proposed that the crucial factor causing a bias lies in the structure of the information: a primacy effect will occur only if the information is correlated over time, and is constantly low. The theory characterizes the correlation of information as category information (CI), and the sensory saliency at each time point as sensory information (SI). They proposed that only in a situation with low SI and high CI will a primacy effect be observed. In our study, we attempt to test this theory in a low-level perceptual context (Exp. 1) and a high-level one (Exp. 2 and 3).

In experiment 1, we tested the theory in a low-level perceptual context. The stimuli are sinusoidal gratings embedded in band-pass-filtered noise (Fig. 1b), oriented in either +45 or -45 degrees. In each trial, subjects see ten frames of the grating, presented in a temporal sequence, and are instructed to choose the predominant orientation. We designed two conditions, one with low CI and the other low SI. The low CI condition is reflected by less consistency of the orientation across different frames within a trial, while the low SI condition is reflected by low contrast of the grating. While CI is low, SI is maintained high and vice versa.

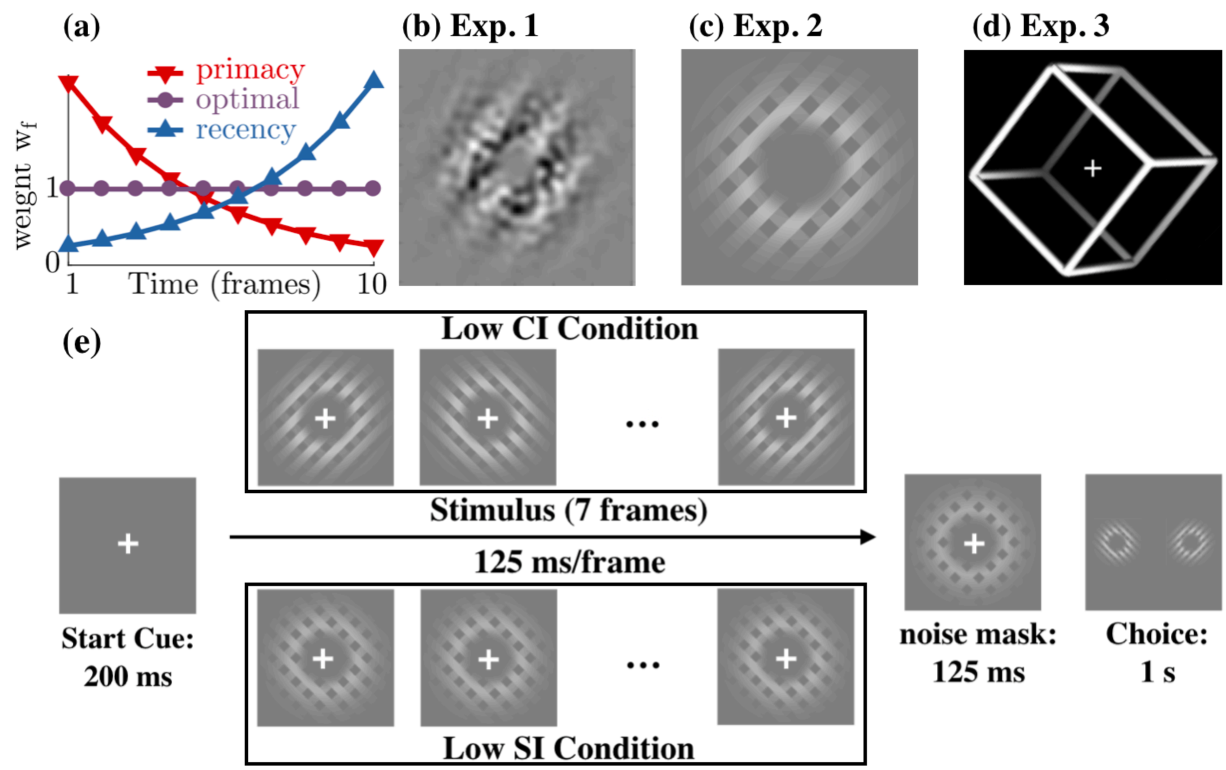

Fig. 1: (a) Three common motifs of temporal integration, shown by the weight decreasing along time (primacy), remaining constant (optimal) or increasing (recency); (b) a right-tilting grating embedded in noise; (c) a right-foreground grating; (d) a Necker cube visual favoring one precept; (e) the flow of one trial of Exp. 2. A start cue precedes 7 frames of stimuli, which are followed by a noise mask, and one second for the subject to choose the foreground orientation.

Results. We did not find a primacy effect in the low SI condition, contrary to the model prediction. The behavioral data is analyzed using logistic regression (see Formula 1). The

$$p(choice|\vec{x}) = \frac{\alpha}{2} + \frac{1-\alpha}{1+e^{-b-\sum_{k=1}^{10} s_k x_k}} \quad (1)$$

weights (denoted by s) reflect how much the information contained in one frame in time influences the final decision in a trial. A larger weight reflects a greater influence of one frame on the final decision. We then plot the weights against time in the course of a trial (Fig. 2). Our prediction is that only in the low SI condition will we see a primacy effect (similar to the red line in Fig. 1a). However, although the original study observed a robust primacy effect in all 12 subjects, our replication yielded near-optimal effect for all 5 subjects (see Fig. 2b, which is similar to the purple line in Fig. 1a). We propose that the lack of a primacy effect may be due to the nature of the noise. We have used stimulus with band-pass-filtered noise added to the grating, which may have caused the stimulus to be resolved differently by the human brain, in a way that the primacy effect did not arise. This explanation proposes a possible relationship between the nature of noise and the interpretation of the human brain that calls for further investigation.

In experiment 2 and 3, we plan to test the same model on a cognitive level that involves abstract processes. We use a foreground-background stimulus in Exp. 2 and a Necker cube in Exp. 3. Exp. 2 asks subjects to segregate the foreground from the background, out of the stimulus consisting of two overlapping gratings (Fig. 1c). In Exp. 3 (Fig. 1d), subjects see a visually bi-stable image, of which the stimulus strengthens one precept of the two possible precepts. For both experiments, we ask the subject to choose the overall image they perceive at the end of a trial. As in Exp. 1, the information contained in a frame within a trial is either high but contradictory (low CI) or constantly low (low SI). An example is shown in Fig. 1e, which demonstrates the procedure of one trial in Exp. 2. The motivation behind the new designs is to test the theory on a high-level domain of sensory perception.

To conclude, the current study aims to test the robustness as well as the generalizability of the theory connecting the primacy effect to the information structure [4]. As the results from Exp. 1 suggest, the presence of a primacy effect may depend on a delicate manipulation of the relationship between the signal and the noise. In the ongoing experiments, we test the same hypothesis on a task in a higher level domain of perception. Whether the theory generalizes to a different domain offers insight on the underlying mechanism of confirmation bias across different levels of sensory perception.

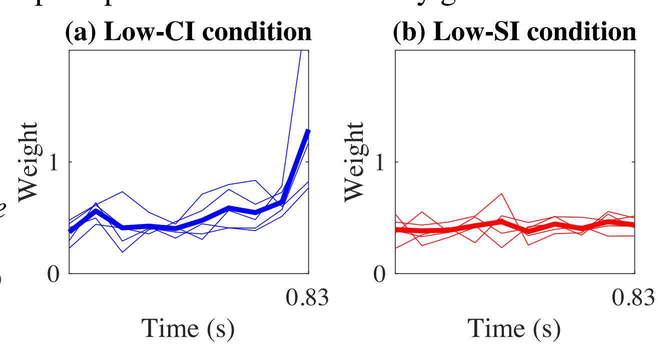

Fig. 2: the regressed weights plotted against time in a trial. The mean of temporal kernels across five subjects is drawn in bold.

Reference:

[1] R. Kiani, T. D. Hanks, M. N. Shadlen, *The Journal of Neuroscience* 28, 3017 (2008). [2] H. Nienborg, B. G. Cumming, *Nature* 459, 89 (2009). [3] B. W. Brunton, M. M. Botvinick, C. D. Brody, *Science* 340, 95 (2013). [4] Lange, R. D., Chattoraj, A., Beck, J., Yates, J., & Haefner, R. *bioRxiv*, 440321. (2018).