Confirmation bias in the temporal integration of evidence

Leslie Li
Mentor: Prof. Ralf Haefner

Abstract

Confirmation bias is prevalent in human behavior, including decision-making (Nickerson, 1998). Specifically, the same amount of evidence may not influence the observer to the same degree when the observer holds a confirmation bias. This phenomenon can be characterized by a primacy effect, that in the course of decision-making, later evidence is weighed less as the observer is biased towards a belief formed from earlier evidence. As a result, in perceptual decision-making tasks in which information is distributed equally over the course of one trial, confirmation bias could be characterized by a decreasing weight of evidence over time. The primacy effect described above is found in some studies (Kiani et al., 2008; Nienborg & Cumming, 2009) but not others (Wyart et al., 2012; Brunton, et al., 2012; Raposo et al., 2014). These studies differ in various aspects, and the question remained unanswered as to why the studies found conflicting patterns of temporal integration of evidence. Recently, a work proposes a theory to explain the discrepancy (Lange et al., 2018). Lange et al. theorized that the type of uncertainty in the stimuli statistics determines the strength of the confirmation bias, and designed a visual perceptual task to test the theory, with results supportive to their theory. The current study seeks to expand on the theory with two experiments. Experiment 1 replicates the perceptual experiment in Lange et al. (2018) on a conceptual level, and Experiment 2 generalizes the theory to a different domain, foreground-background segmentation. Results of Experiment 1 found no primacy effect, contradictory to the theory prediction, while results of Experiment 2 showed a primacy effect as predicted by the theory. The current study both expands and poses limitation to the theory of confirmation bias, and calls for further investigation.

Confirmation bias in the temporal integration of evidence

Introduction

Confirmation bias occurs everywhere in human behavior, including decision-making (Nickerson, 1998). In a simple scenario of drawing balls from a bag, if the observer has a belief that the bag contains more red balls than blue balls, seeing yet another red ball will easily confirm such a belief, while seeing a blue ball might be dismissed as an exception. In a sense, the same amount of evidence may not influence the observer to the same degree when the observer forms a confirmation bias. This phenomenon can be characterized by a primacy effect, that in the course of decision-making, later evidence is weighed less as the observer is biased towards a belief formed from earlier evidence. As a result, in perceptual decision-making tasks in which information is distributed equally over the course of one trial, confirmation bias could be characterized by a decreasing weight of evidence over time.

The primacy effect described above is found in some studies (Kiani, Hanks, & Shadlen, 2008; Nienborg & Cumming, 2009). However, other studies have found that information is weighted equally over time (Wyart, De Gardelle, Scholl, & Summerfield, 2012; Brunton, Botvinick, & Brody, 2012; Raposo, Kaufman, & Churchland, 2014). As an example of primacy effect, Nienborg & Cumming (2009), looking at the neuron-level response of macaque monkeys, designed a visual binocular disparity task, and the results suggested a bias towards the earlier trials. On the other hand, Brunton et al. (2012) explored decision-making in both rats and humans in an auditory Poisson click task, finding no bias in temporal weighting (i.e., optimal integration of evidence). These studies differ in the test modality, task design, species, and methodology, and the question remained unanswered as to why the studies found conflicting patterns regarding temporal integration of evidence.

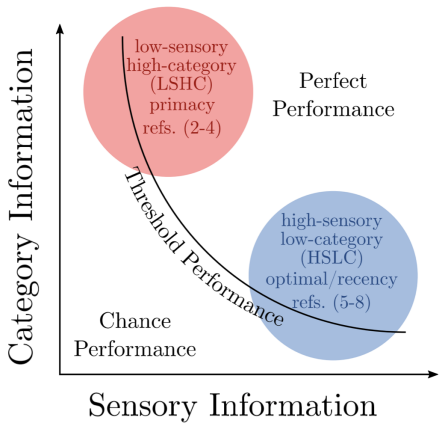

Figure 1: The space describing the information structure of the perceptual decision tasks. The theory predicts a primacy effect in the red region, when SI is low and CI is high. Figure adapted from Lange et al. (2018).

Recently, a work proposes a theory to explain the discrepancy in the literature (Lange, Chattoraj, Beck, Yates, & Haefner, 2018). Lange et al. theorized that the type of uncertainty in the stimuli statistics determines the strength of the confirmation bias, and

designed a visual perceptual task to test the theory, with supportive results. The theory measures the perceptual decision-making tasks in two dimensions, sensory information (SI) and category information (CI, see Figure 1). SI represents the sensory salience at each time point: for example, if a visual stimulus is fainter, SI would be low; if it is clearly visible, SI would be strong. CI represents the consistency of information: CI is low when information conflict over time, and high if information remains consistent. When both SI and CI are high, the task will be easy and performance will be at ceiling; when both SI and CI are low, the task will be hard, and performance will be at chance. The critical prediction of the theory is that when SI is low and CI is high, confirmation bias will be strong; as a comparison, when CI is low and SI is high, a confirmation bias will not show. Following this theory prediction, Lange et al. (2018) designed a visual task with band-pass filtered white noise as stimuli, with one condition of low SI and another of low CI. Results indicated that a primacy effect is observed only in the low SI condition, consistent with theory prediction.

The current study seeks to expand on the theory with two experiments. Experiment 1 follows the design of Lange et al. (2018) and attempts to replicate the perceptual experiment in Lange et al. (2018) on a conceptual level, and Experiment 2 generalizes the theory to a different task, foreground-background segmentation. The rest of this paper will describe Experiment 1 and 2 in order, followed by a general discussion of the results.

Experiment 1

Method

Four naïve participants and an informed author participated in this study. All participants are undergraduate students from the University of Rochester. Subjects enrolled all reported to have corrected-to-normal vision. Subjects who fail to return and complete all five sessions, or who cannot complete/have confusion about the task would be discarded (none in this experiment).[1] Each subject is compensated with $10/hr.

This experiment uses a within-subject design across two conditions, the low SI condition and the low CI condition. The flow of a trial and examples of the stimuli in either condition are shown in Figure 2. The low SI condition takes 4 sessions, while the low CI condition takes 2 sessions. Each session contains eight 100-trial blocks. The low SI condition takes more trials because more data is required for the statistical analysis, as the sensory signal is faint.

As shown in Figure 2, Each trial contains ten frames of visual stimulus. Each stimulus consists of a grating (sinusoidal wave) tilted either $-45°$ (left) or $+45°$ (right), and additive Gaussian noise. The noise is generated using the same method as Lange et al. (2018), using a band-pass-filtered white noise with zero signal. A mask in the shape of an annulus is applied to the stimulus, resulting in an image as shown in Figure 1. During the experiment, a small cross is present in the middle of the grating to be fixated on. Stimuli in each trial are preceded by a "start" cue and followed by a noise mask. After each trial, a left-tilted and a right-tilted grating is shown on the screen, and the subject is asked to

[1] Other exclusion criteria are that (1) subject demonstrates a lack of attention, characterized by $\alpha > 0.1$, (2) subject demonstrates a strong bias towards one choice, characterized by $\mid b - 0.5 \mid > 0.2$. For the meaning of these terms, see the Statistical Analysis section.

press either “1” or “3” on a numeric keypad to respond “left” or “right”– the direction which the preceding stimuli are “most consistent with”. Subjects receive audio feedback that indicates a correct choice (high-pitch beep), an incorrect choice (low-pitch beep), or no response within the allowed response window (three low beeps).

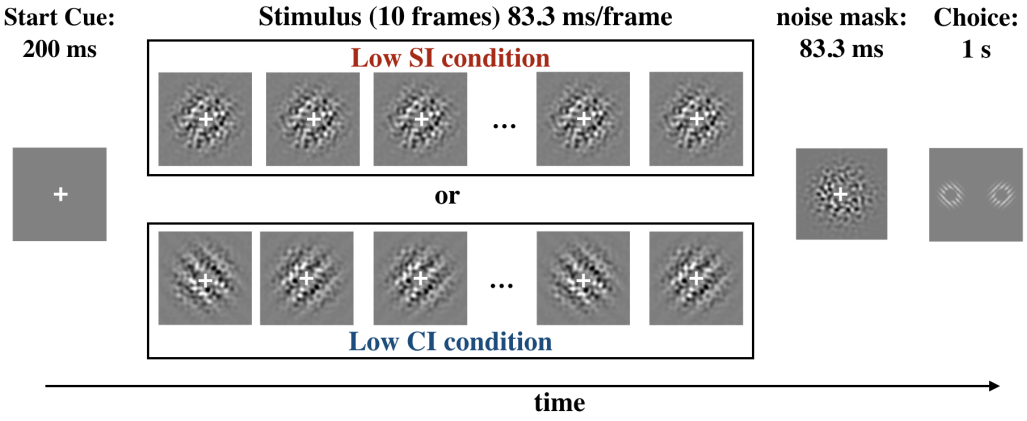

Figure 2: Demonstration of the flow of one trial in Experiment 1.

The independent variable (signal strength) in each trial is determined by a 2:1 staircase, in which difficulty increases after every two correct trials, but decreases with every incorrect trial. For the low SI condition, the contrast of the grating decreases as the subject gets trials correct. This is done while the noise level and left-right frame ratio are kept constant (the ratio is always kept as 1:0, such that there are no inconsistent frames). For the low CI condition, the left-right ratio of the ten frames approaches 0.5 as trials become more difficult, while contrast is kept high. The staircase resets at the beginning of each block. Trials always begin with high contrast and consistent ratio, and depending on the condition, one of the two independent variables is manipulated by the staircase.

At the beginning of the experiment, instruction as well as two tilted sample gratings are displayed on the screen. Meanwhile, the researcher reads the instruction to the subject, including the task, experiment duration, and response buttons. The subject then starts the experiment. After each session, the researcher will check in on the subject and ask if they want to take a break. The number of completed blocks will be displayed on the screen during the break. During the experiment, subjects can quit any time by pressing the backspace key. The mean duration for each 8-block session is one hour, and each subject performs no more than one session per day to avoid fatigue.

Statistical Analysis

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

The behavioral data is analyzed using logistic regression (see Formula 1). The signal contained in each frame of stimulus is denoted by the vector $\vec{x}$, while x_k denotes the signal of each frame. The signal is obtained by calculating the signal strength in $-45°$ and $+45°$ in the Fourier domain. α denotes the lapse rate that accounts for subject's unattended guessing (Wichmann & Hill, 2001); b denotes the bias in the subjects' choices; $\vec{s}$ denotes

the weights, one for each frame, which reflect how much the subjects' decision at the end of a trial is influenced by the signal (x_k) of a particular frame k. A larger weight reflects a greater influence of one frame on the final decision. A smoothing hyper-parameter is also introduced into the model to punishes the second-order derivative of the weights (i.e., the rate of change of the weights). This prevents the weights from drastically changing locally, and is motivated by our prior belief that the weighting of evidence should not fluctuate drastically within tens of milliseconds.

The above model-fitting process is repeated 100 times for each condition of each subject, using the bootstrapping method on the data to obtain error estimates. After the regression, we report the medians of the weights. We obtain the slope of a linear fit to each of kernels, and use the slope value to test our hypothesis – a negative slope would be in favor of our hypothesis of a primacy effect, and a flat kernel (i.e., zero slope) or positive slope would indicate the lack of a primacy effect.

Results

The results suggest that subjects do not exhibit any primacy effect, contrary to the theory prediction. First, psychometric functions of each subject are calculated for both conditions, to examine the performance of individual subjects (see Figure 3). None of the 5 subjects are excluded based on the exclusion criteria (see the Method section), and the author data do not differ from the naive subjects.

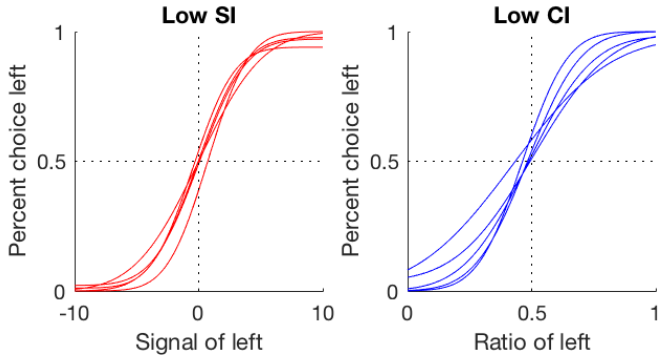

Figure 3: Psychometric functions of the subject response in Experiment 1.

Then, via logistic regression reported in the previous section, temporal kernels for each subject and each condition is obtained. As shown in Figure 4, the Low SI conditions yielded temporal kernels that were flat, contrary to the prediction of a decreasing kernel. In the Low CI condition, the temporal kernels were also mostly flat, except for the last frame, where evidence had a large effect. We attribute that to an error in the design, that the noise mask is not strong enough to mask the effect. As a result, the visual aftereffect allowed the last frame to have a large weight in the subjects' decision. We have modified the design and used a stronger noise mask in the following experiment, and this effect did not show. The critical inference drawn from the results of Experiment 1 is that subjects did not show a primacy effect that the theory predicts.

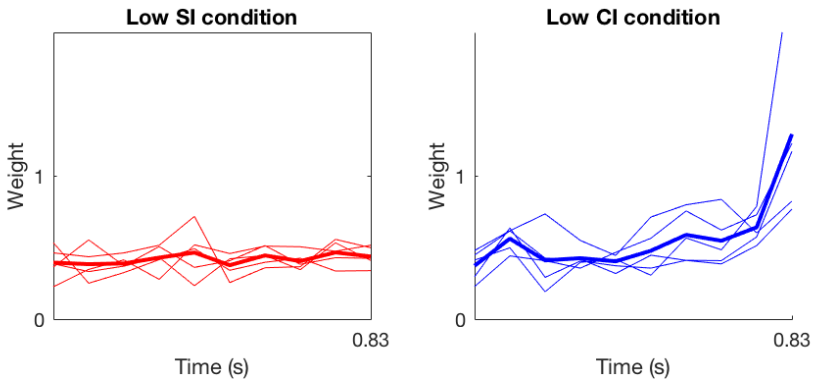

Figure 4: The temporal kernels of all five subjects in Experiment 1. Each line represents one individual subject, and the thick line represents the mean across all subjects.

Experiment 2

Method

Experiment 2 follows the design of Experiment 1, except for a different set of stimulus. A demonstration of the stimulus and design is shown in Figure 5. In Experiment 2, square waves replace the sinusoidal waves in the stimuli, and instead of only showing one orientation, both $-45°$ and $+45°$ are shown in every frame. However, one of the gratings will be obstructing another. The task for the subjects is to report the orientation of the grating that is on the top (i.e., in the foreground). This design tests subjects on the domain of foreground-background segmentation, which is a common motif in visual processing.

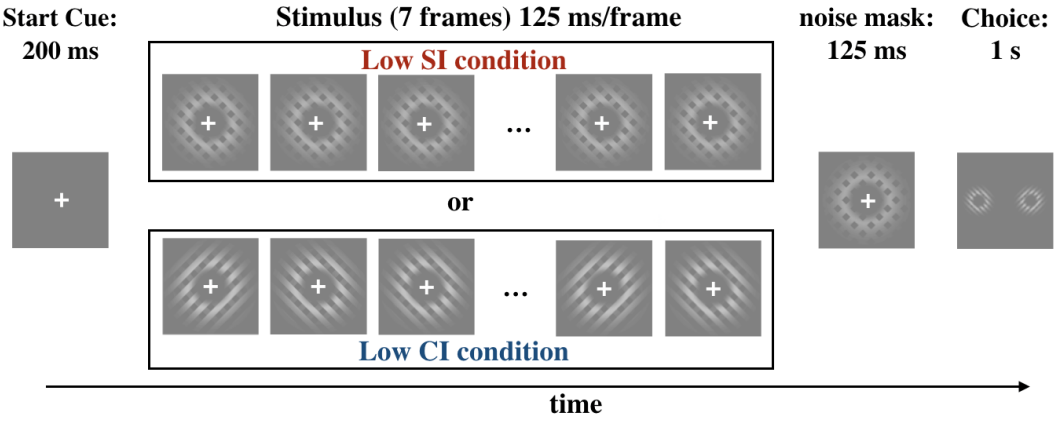

Figure 5: Demonstration of the flow of one trial in Experiment 2.

The Low SI condition in Experiment 2 is conducted in two versions, 2A and 2B. While the phase of gratings is randomized across frames in both versions, the critical difference between 2A and 2B is that 2A did not use complete randomization. This leads to the possibility of the subjects, due to a confirmation bias, actively making eye movements to the part of the image that is consistent with their beliefs. Experiment 2B controls for the active searching by making the phase randomization of each frame completely independent of that of the other frames. This makes sure that the primacy effect, if observed, is not due to the different input to the brain (i.e., external) but arises from the internal computation of the brain.

Another methodological difference in Experiment 2 is that the number of frames is reduced from 10 to 7, and the time duration between two frames is extended from 83.3ms to 125ms. This is due to the observation from a pilot version of the experiment, that subjects report needing slower presentation of stimuli such that they have more time to process the information. In each of the 125ms-long frame, the image is shown in only the first 83.3ms, and the remainder of the time is filled by the grey background and the fixation sign. The blank frame is added in order to prevent the sensation that the image is moving, due to the phase randomization described above.

For each subject, 2 sessions are collected for each Low SI condition, and one session is collected for the Low CI condition. Due to time constraints and subject availability, not all subjects have finished all the two sessions for Experiment 2B, and not all subjects have been available to return for the Low CI condition. Here, all the data collected are reported except for subject excluded due to inability to perform the task (one out of 10 subjects in Experiment 2A). As of this report, 9 subjects have participated in Experiment 2A, 7 in Experiment 2B, and 4 in the Low CI condition of Experiment 2.

Statistical Analysis

The signal is calculated from the difference of the luminance value of the two orientations, with -0.5 corresponding to a right-tilted grating only, $+0.5$ corresponding to a left-tilted grating only, and 0 corresponding to both gratings of equal luminance values, where there is no evidence that either grating is on the top of the other. The rest of the logistic regression and bootstrapping are identical to those of Experiment 1.

Results

The overall results from Experiment 2 show a primacy effect in both Low SI conditions, and a comparison to the Low CI condition suggests that the primacy effect arises from the different information structures in the perceptual task. Following Experiment 1, we first calculate the psychometric function of each subject for each condition (Figure 6). All subject except one passed the criteria for exclusion: In experiment 2B, one subject (subject 05) has a lapse rate of >0.1. The data for this subject for this condition is therefore removed in the following analysis. Individual data for this subject is shown in black in Figure 6 and 7 to identify the excluded data.

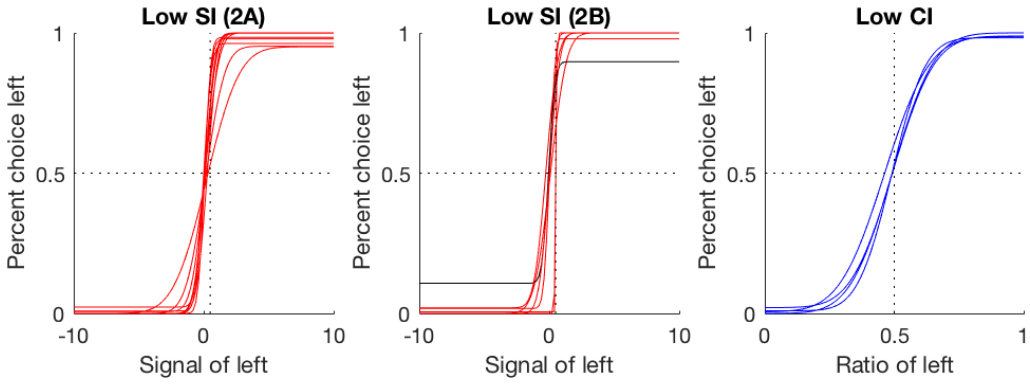

Figure 6: Psychometric functions of the subject responses in Experiment 2.

Then, the temporal kernel for all subjects are obtained via logistic regression (Figure 7). Results show a primacy effect in both Experiment 2A and 2B, as the overall weight decrease over time. The Low CI condition, in comparison, did not show a primacy effect, suggesting that the primacy effect is due to the low SI design.

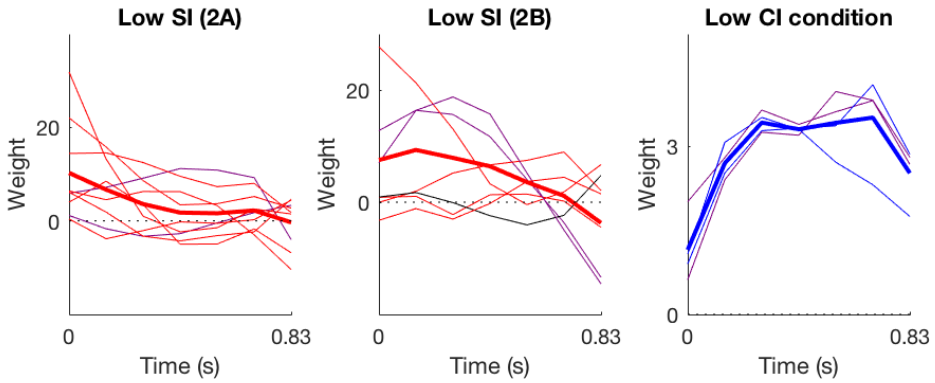

Figure 7: The temporal kernels of all subjects in Experiment 2. Each line represents one individual subject, and the thick line represents the mean across all subjects. Purple lines represent author data.

In Figure 7, it can be observed that although the author data (colored in purple) do not differ significantly from those of the naive subjects in Experiment 2A and the Low CI condition, in Experiment 2B, both authors showed sharply decreasing kernels compared with the average performance. It is therefore possible that the knowledge about the experiment has influence over the data. However, it is also notable that many naive subjects in Experiment 2B have not finished 2 sessions, and that it is possible that logistic regression failed to extract a reliable kernel based on the small amount of data of these subjects. Therefore, although primacy effect is observed in Experiment 2B, more data is needed to draw a more conclusive inference.

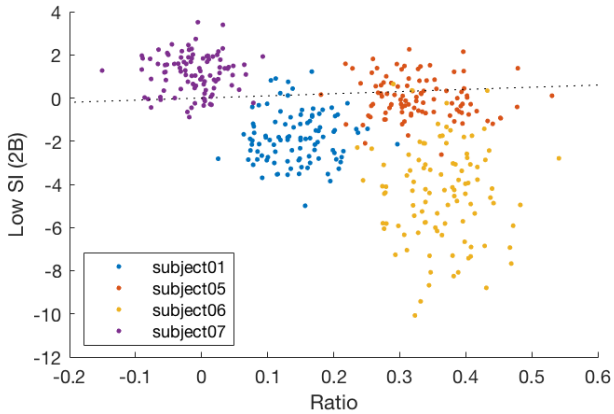

Figure 8: Bootstrapped slopes of temporal kernels. Dotted lines denote the identity line.

Lastly, in order to compare the performance across conditions, we take the 100 bootstrapped samples from the Low SI condition (from Experiment 2B) and the Low CI condition, and fit a linear regression to each temporal kernel obtained from the

bootstrapped samples. The analysis yields 100 linear fits for each condition for each subject. We then compare the slopes of these linear fits. The critical prediction is that the Low SI condition, due to the primacy effect, would have a more negative slope than the Low CI condition. Results are shown in Figure 8. For 2 of the 4 subjects (subject 01 and 06) who gathered enough data for this comparison, the result is in accordance with the theory prediction, that the distribution of the bootstrapped results lies below the identity line. Whereas this is not the case for subject 05 and 07, it should be remembered that the data of subject 05 is excluded due to a high lapse rate, and should not be taken into consideration in the analysis. The figure only includes the results of subject 05 for completion. While this within-subject comparison only gives a weak evidence in favor of the theory prediction, it should also be noted that this analysis only includes a small portion of the subjects.

Discussion

Experiment 1 is a conceptual replication of Lange et al. (2018), as a test to their theory of confirmation bias. It uses the same design except for some change in the visual stimulus. However, the results of Experiment 1 do not show a confirmation bias. On the other hand, Experiment 2 extends the theory of confirmation bias by testing it in a foreground-background segmentation task. The results of Experiment 2 are consistent with theory prediction, in that confirmation bias, manifested in a primacy effect, is observed when and only when SI is low.

Whereas the results of Experiment 2 can be explained well by the theory predictions, the theory does not explain the lack of a primacy effect in Experiment 1. Since the primacy effect was robust in the behavioral experiment from Lange et al. (2018), in that all 10 subjects showed a confirmation bias, it is unlikely that the lack of primacy effect in all five subjects in Experiment 1 is due to chance. Below, we offer a possible explanation for the results obtained in Experiment 1. Lange et al. (2018) used a band-pass filtered noise that is filtered along the orientation of $-45°$ or $+45°$ in Fourier space. This leads to oval-like bands tilted in either direction to be observed by the subjects. The two orientations cannot be both present at the same time, as that would lead to star-shaped objects that are never observed through the experiment. As a result, subjects can maintain a belief that a $-45°$ and a $+45°$ choice cannot both be true. Similarly, in Experiment 2, the foreground-background segmentation task defines by nature that one grating must be on top of each other. Therefore, the two choices "left" and "right" compete against each other during the course of evidence accumulation. That is, evidence in favor of "left" will not only strengthen the decision towards "left", but also inhibit the decision against "right." Such a competitive nature of the two choices is, however, absent in the design of Experiment 1. The visual stimulus is composed of superimposition of a sinusoidal wave and a noise image. As a result, as subjects try to detect a sinusoidal wave of either direction from random noise, evidence in favor of a "left" choice does not make the "right" choice less possible. In other words, the two choices "left" and "right" do not compete with each other in Experiment 1.

Tentatively, we hypothesize that in order to trigger a confirmation bias, the information structure should not only be low in SI and high in CI (as Lange et al.

predicted), but the candidates of the decision should be competing with each other. To illustrate this hypothesis in another setting, consider the binocular disparity task in Nienborg & Cumming (2009). As Lange et al. (2018) explained, the task is low in SI in that the stimulus is very close to the reference plane, and thus evidence is only weakly predictive of the true answer; CI in the task is high in that evidence is consistent over time. However, based on our new hypothesis, it should be added that the stimulus can either be closer or farther compared with the reference plane, it cannot be both. As a result, the two choices are competitive with respect to each other.

This hypothesis, although tentative, offers some future directions of interest. For example, in both the visual and the auditory domain, there exist stimuli that have strongly binary interpretations. An example from the visual domain could be the Necker cube. As a visually bi-stable image, it can be interpreted in two different ways, but both interpretations cannot be true at the same time. In the auditory domain, an example would be phoneme perception: while a sound can be interpreted as a /b/ or a /p/ in different contexts, a sound cannot be /b/ and /p/ at the same time. These new directions are worth exploring. On the other hand, work is in progress on building a computational model with separate integration of evidence for two choices (i.e., "left" and "right" choices do not inhibit each other). If such a model does not exhibit a confirmation bias in a low SI, high CI setting, the hypothesis proposed above would be supported.

In conclusion, the current work explores and expands on Lange et al. (2018). While we successfully generalized the theory to another task domain (foreground-background segmentation), a task that closely resembles the original task design in Lange et al. did not find a primacy effect. We attribute that to the nature of the stimulus that was yet to be defined by the theory, and call for future research.

Acknowledgement

I thank Dr. Ralf Haefner as the mentor of this project and Ankani Chattoraj for help throughout the project. I also thank Dr. Chigusa Kurumada as the instructor of BCS 206–207, and my classmates for their support and feedback.

Reference

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science, 340*(6128), 95-98.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience, 28*(12), 3017-3029.

Lange, R. D., Chattoraj, A., Beck, J., Yates, J., & Haefner, R. (2018). A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *bioRxiv*, 440321.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220. doi:10.1037/1089-2680.2.2.175

Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons

reflects more than a neuron's causal effect. *Nature, 459*(7243), 89-92. doi:10.1038/nature07821

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature neuroscience, 17*(12), 1784.

Wichmann, F. A., Carlo simulations, M., & Jeremy Hill, N. (2001). *The Psychometric Function I: Fitting, sampling, and goodness-of-fit.* Retrieved from http://www.psy.gla.ac.uk/ martinl/Assets/MCMPS/Wichmann&Hill01a.pdf

Wyart, V., De Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron, 76*(4), 847-858.