Relating confidence judgements to temporal biases in perceptual decision-making

Ankani Chattoraj*, Martynas Snarskis, Ralf M. Haefner*

Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA.
*Correspondence: achattor@ur.rochester.edu, ralf.haefner@gmail.com.

Abstract

Decision-making is often hierarchical and approximate in nature: decisions are not being made based on actual observations, but on intermediate variables that themselves have to be inferred. Recently, we showed that during sequential perceptual decision-making, those conditions induce characteristic temporal biases that depend on the balance of sensory and category information present in the stimulus. Here, we show that the same model makes predictions for when observers will be over-confidence and when they will be underconfident with respect to a Bayesian observer. We tested these predictions by collecting new data in a dual-report decision-making task. We found that for most participants the bias in confidence judgments changed in the predicted direction for stimulus changes that led them from an over-weighting of early evidence to an equal weighting of evidence or an over-weighting of late evidence. Our results suggest that approximate hierarchical inference might provide the computational basis for biases beyond low-level perceptual decision-making, including those affecting higher level cognitive functions like confidence judgements.

Keywords: perceptual decision-making; approximate inference; feedback; choice bias; confidence judgements

1 Introduction

Humans make decisions by integrating information over time. Previous work has identified temporal biases during perceptual decision-making in which participants' decisions are influenced more by early evidence (primacy effect; (Kiani et al., 2008; Nienborg and Cumming, 2009)), later evidence (recency effect; (Drugowitsch et al., 2016)), or weighted equally across the trial (optimality; (Wyart et al., 2012; Brunton et al., 2013)) in situations when optimal decision-making demands equal weight given to all pieces of evidence (Figure 1A).

In a recent study, we proposed a new theory for these biases (Lange et al., 2020). The theory first acknowledges that the brain does not use sensory observations (e.g. activity in the retina) directly to make decisions, but instead bases them on intermediate sensory features (e.g. in visual cortex). This implies a partition on the information each stimulus holds about the correct choice. "Sensory information" represents the amount of information between stimulus and intermediate feature. "Category information" represents the information between feature and correct choice. For sequential tasks, consisting of multiple stimulus frames, this typically corresponds to the proportion of frames which are consistent with the correct choice. In the approximate hierarchical inference model of our previous work (Lange et al., 2020), the temporal bias depends on the balance of sensory and category information such that the primacy effect is seen when category information dominates while the recency effect and optimal weighing are seen when sensory information dominates (Figure 1A) – in agreement with prior empirical studies (Figure 1C).

In this study, we investigated whether the sensory information and category information would also affect another commonly studied behavior: confidence. Confidence is usually defined as the belief of a participant that their choice in a task was correct (Pouget et al., 2016; Grimaldi et al., 2015; Meyniel et al., 2015; Li and Ma, 2020). Confidence judgements are known to be systematically influenced by certain stimulus statistics including volatility, whose effect has been linked to evidence integration (Zylberberg et al., 2016; Castañón et al., 2019). In these studies, increasing

volatility of the stimulus made participants more confident (despite similar accuracy).

Investigating our hierarchical approximate inference model from previous work, we found that a primacy bias should entail over-confidence, and that a recency bias should entail under-confidence compared to a Bayesian observer. We next tested these predictions by collecting new data from a perceptual discrimination and confidence judgment task. We both replicated our previous results on temporal weighting biases, and found that stimulus conditions which induced a primacy bias in participants also made them overconfident, compared to stimulus conditions that induced a flat weighting or recency bias.

2 Hierarchical Approximate Inference

We follow (Lange et al., 2020) in implementing a sampling-based approximate inference model of the visual discrimination task in Figure 1D-E (see also Visual Discrimination Task section). In this task, participants determine whether a series of noisy oriented stimuli had more frames tilted left or right. The ideal observer chooses the most probable choice $C \in \{-1, +1\}$ by integrating the sensory evidence e_f in each frame f over F independent frames according to Bayes rule:

$$p(C|e_1, \ldots, e_F) \propto p(C) \prod_{f=1}^{F} p(e_f|C)$$

Since decision-making areas in the brain do not have access to peripheral sensory observations, the inferred intermediate sensory representation x needs to be integrated out (see Figure 1B). This can be done according to the Sequential Probability Ratio Test (Gold and Shadlen, 2007), wherein a running estimate of the log posterior odds LPO_f is updated every frame:

$$\begin{aligned}
&\underbrace{\log \frac{p_f(C=+1)}{p_f(C=-1)}}_{\mathrm{LPO}_f} \equiv \log \frac{p(C=+1|e_1,\ldots,e_f)}{p(C=-1|e_1,\ldots,e_f)} \\
&= \log \frac{p_{f-1}(C=+1)}{p_{f-1}(C=-1)} + \log \frac{p(e_f|C=+1)}{p(e_f|C=-1)} \\
&= \underbrace{\log \frac{p_{f-1}(C=+1)}{p_{f-1}(C=-1)}}_{\mathrm{LPO}_{f-1}} + \underbrace{\log \frac{\int_x p(e_f|x)p(x|C=+1)\mathrm{d}x}{\int_x p(e_f|x)p(x|C=-1)\mathrm{d}x}}_{\hat{\mathrm{LLO}}_f} \\
&= \mathrm{LPO}_{f-1} + \hat{\mathrm{LLO}}_f
\end{aligned} \tag{1}$$

where $\hat{\mathrm{LLO}}_f$ is the estimate of the log-likelihood ratio implied by evidence e_f via sensory intermediate x.

Exact inference in this model produces an equal weighting of evidence (see Figure 2F). Temporal bias occurs under the assumptions that (1) the intermediate sensory representation x does not strictly encode the choice likelihood for that frame, but is modulated by top-down feedback connections which act like a prior on x, incorporating current beliefs about stimulus category (Figure 1B) and (2) inference in the model is *approximate* (see (Lange et al., 2020) for more details). The degree to which the prior's influence on the intermediate representation is over- or under-corrected during online processing determines the temporal weighting within a trial. Specifically, under-correcting for the influence of the prior leads to an effective double-counting of category beliefs and subsequent primacy effect, whereas as over-correcting for the prior effectively leads to forgetting, or leaky integration, and a subsequent recency effect. Importantly, whether the prior's influence is under- or over-corrected is determined by the amount of category information in the stimulus (Figure 1B+C, (Lange et al., 2020) for more details). We will therefore focus on the same two stimulus conditions in our work: Low Sensory & High Category (LSHC) information, and High Sensory & Low Category (HSLC) information — both matched for overall (threshold) performance on the task. In the context of a orientation discrimination task, Figure 1D+E shows the corresponding stimuli.

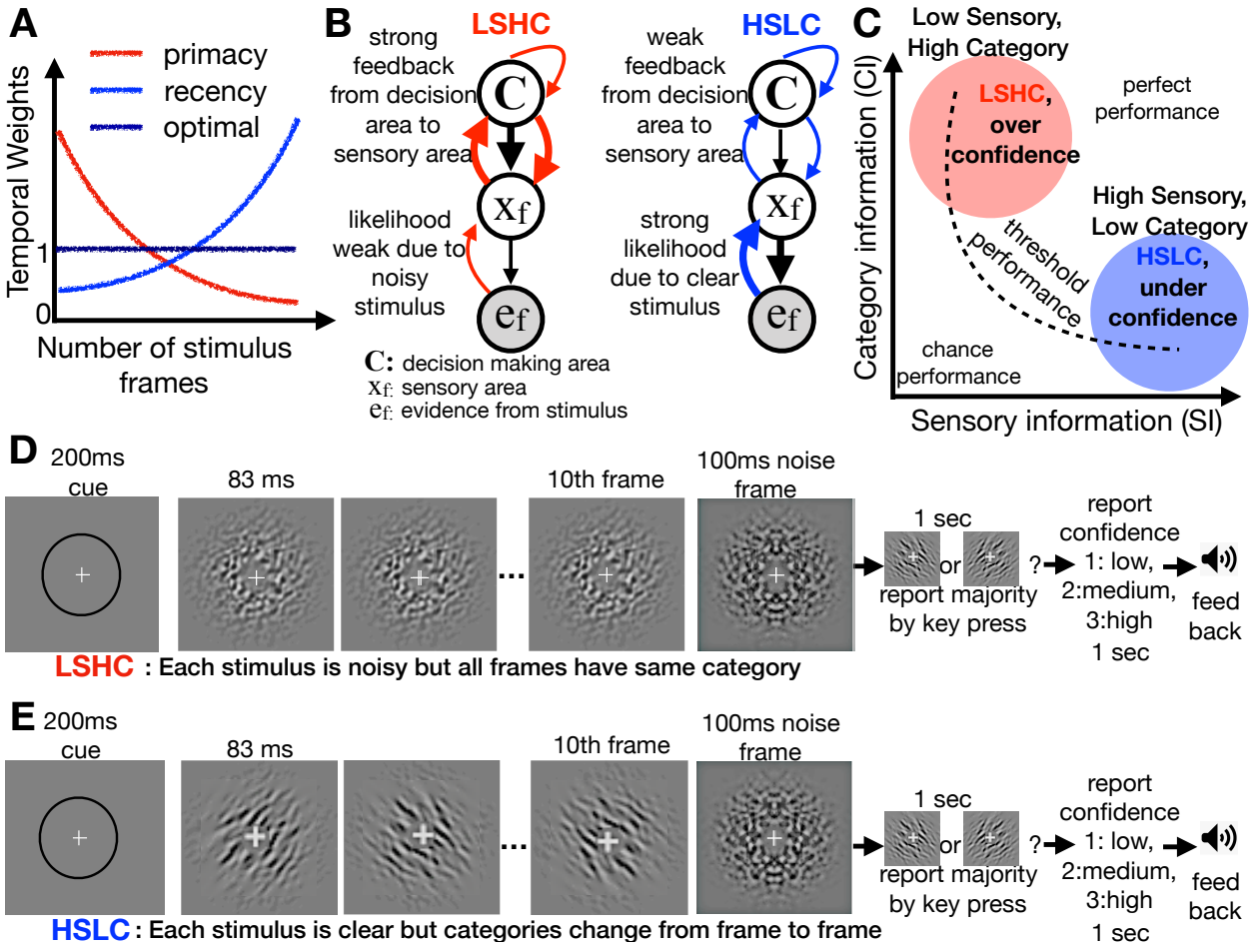

Figure 1: Model and task to test confidence judgements. (A) Example temporal weighting profiles, showing primacy effect (red), optimal weighting (violet), and recency effect (blue). (B) Model proposed by (Lange et al., 2020) for LSHC and HSLC conditions. Evidence in frame e_f is used to infer sensory representation x_f, which is used to update posterior for choice C. LSHC: low sensory information means posterior update will be dominated by prior, which is posterior after last frame, inducing confirmation bias. HSLC: high sensory information means likelihood dominates update, thwarting confirmatory feedback. (C) Task space proposed by (Lange et al., 2020) measuring category (y-axis) and sensory (x-axis) information. Red zone (LSHC) indicates tasks that show primacy effect; blue area (HSLC) indicates tasks that show optimal or recency effect in weighting profile. Our results will show differences in confidence judgements along the same stimulus dimensions (D-E) Two conditions of experimental task. Participant sees 10 frames of filtered noise oriented ± 45 deg and reports the majority orientation. They report their confidence before receiving feedback. (D) LSHC stimulus contains low sensory (orientation hard to detect) but high category (each frame consistent with answer) information. (E) HSLC stimulus contains high sensory (orientation easy to detect) information but low category information (contains inconsistent frames). (A), (B) and (C) have been adapted from (Lange et al., 2020).

2.1 Log posterior odds as proxy for confidence

We assume that reported confidence is monotonically related to the strength of evidence collected throughout a trial, so we use the absolute value of the log posterior odds after the final frame, $|\mathrm{LPO}_F|$ as a proxy for confidence. The model's choice is determined by the sign of LPO_F, while the confidence by its magnitude.

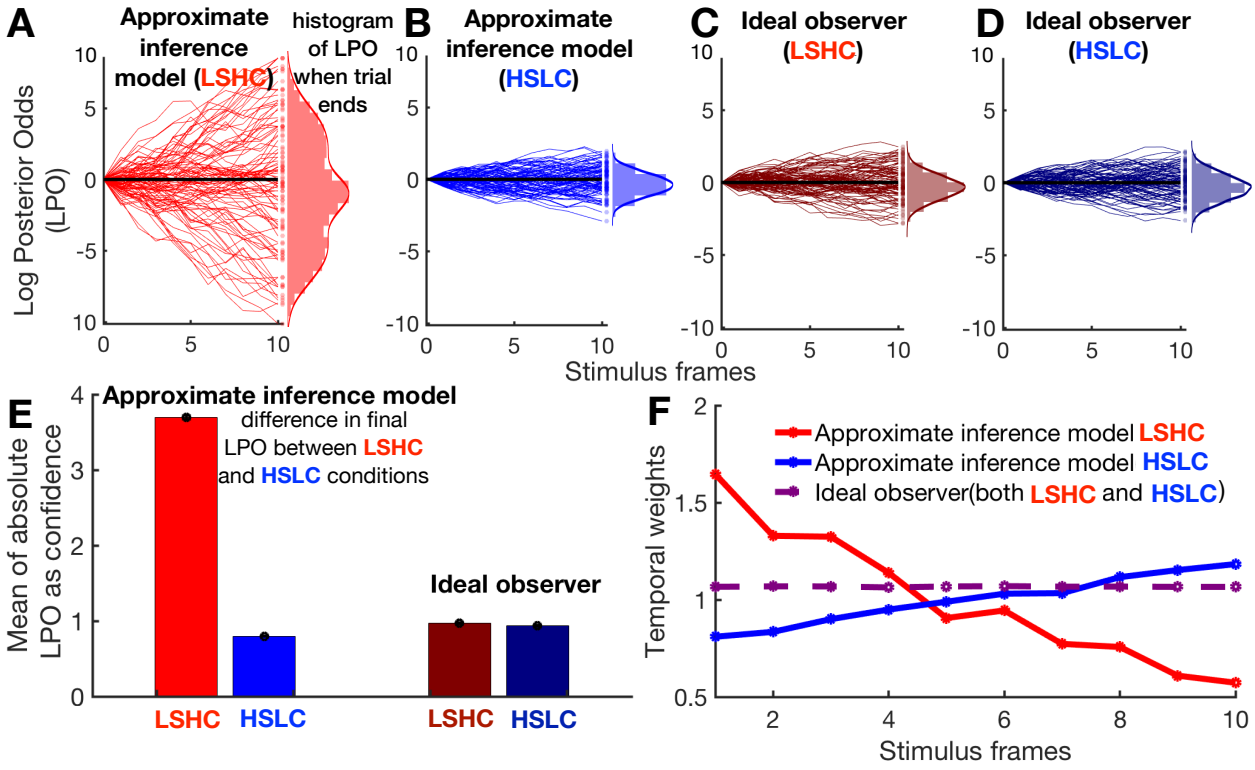

Figure 2: Model simulations for confidence and temporal bias for approximate and idea observers. (A-D) Simulation traces show LPO (y-axis) over 10 frames (x-axis) of LSHC (A,C) stimulus and HSLC (B,D) stimulus for approximate (A,B) and exact (C,D) inference. Absolute LPO at the final frame is taken as a proxy to confidence judgements, whose distribution is shown to the right. (E) Average absolute LPO shown for both approximate and ideal observers and both stimulus conditions. Note LSHC for approximate observer shows much larger confidence, corresponding to the larger LPO spread in (A). (F) Normalized temporal weights for the same simulations. Note ideal observer shows flat, or optimal, weighing profile, while approximate observer shows primacy effect for LSHC and recency effect for HSLC. Thus the model predicts increased confidence to co-occur with primacy bias.

2.2 Feedback between decision area and sensory area causes primacy and high confidence

Using identical model parameters to (Lange et al., 2020), we first confirmed that our hierarchical inference model produced the same temporal biases during approximate, but not exact, inference, as previously reported (Figure 2F). We next investigated the confidence judgements implied by the approximate model and found that they indeed differed between the two stimulus conditions: for the LSHC condition, the final absolute LPO were substantially larger on average than in the HSLC condition. Furthermore, in the LSHC condition, confidence was higher than for an ideal observer (exact inference), and in the HSLC condition it was lower (Figure 2E). The reason for this is intuitive. The same confirmatory (positive feedback) dynamics between the decision-making representation and the sensory representation that lead to an overweighting of early evidence, pull the accumulated log odds *away* from 0, such that their final distribution at the end of the trial is wider than without those feedback dynamics. As a result, the average LPO at the end of the trial is larger than for the ideal observer (Figure 2A). In the absence of this confirmation bias dynamic, the final LPO will agree with those of an ideal observer, and match it in its confidence judgements. The reason that our model produces slightly less variability in the final LPO for HSLC stimuli lies in the fact that our model also contains a small leak (forgetting) term to match the model parameters fit to empirical subject data in (Lange et al., 2020). This leak term pulls LPOs towards zero, and entails a slightly smaller variability in LPO, and hence a smaller average confidence than the ideal observer (Figure 2B). Figures 2C-F confirm that confidence is matched and temporal biases are absence for the ideal observer.

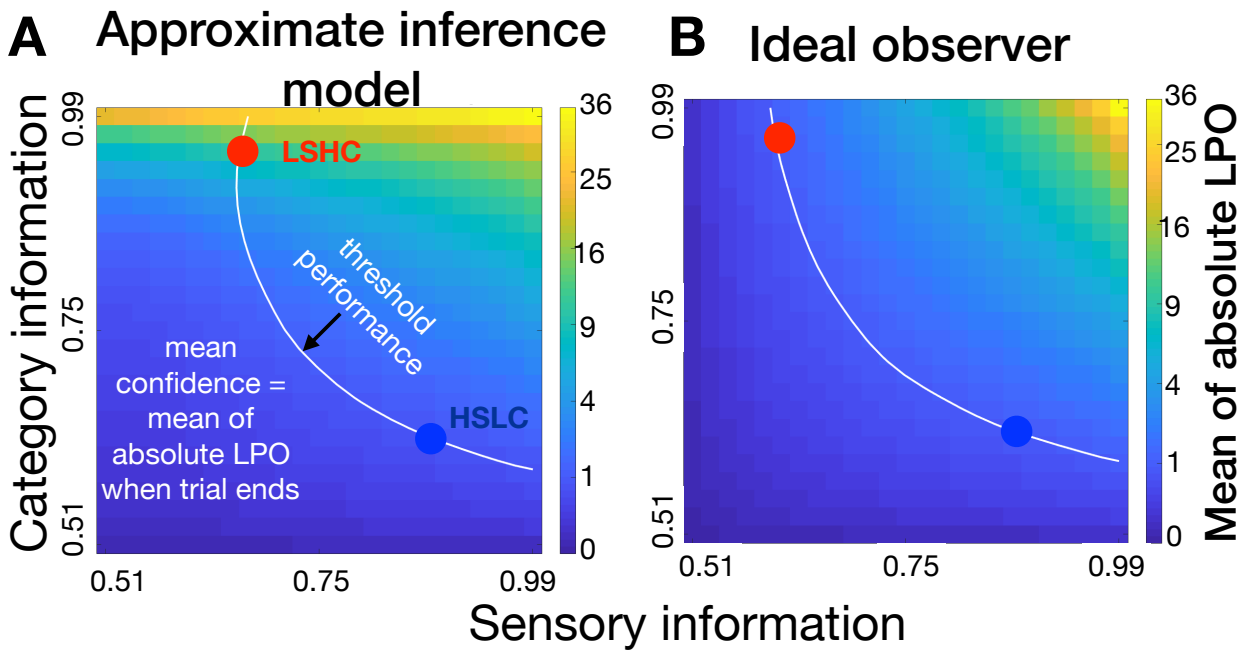

Figure 3: Expected confidence as a function of sensory and category information. (A) Simulations for approximate inference model show confidence grows primarily with category information. (B) Simulations for ideal observer show confidence is equally affected by category and sensory information.

Finally, we characterized the predicted confidence judgements (mean final LPO) for a wide range of sensory and category information values (Figure 3). As expected, simulations of the ideal model, representing a control, show unbiased confidence values that depend equally on sensory and category information (Figure 3B). For the approximate model, on the other hand, confidence values more strongly depend on category information (Figure 3A). This is consistent with the dynamic described in (Lange et al., 2020), that strong beliefs about the category of a trial reinforce themselves by biasing intermediate sensory representations when those beliefs are indeed predictive of future sensory inputs within a trial (high category information). The region of high confidence in the upper part of the plot, is the same region where the primacy bias in the weighting of evidence is the strongest (cf. Figures 4e+h in (Lange et al., 2020)).

3 Visual Discrimination Task

3.1 Rationale

We next tested our model predictions using a dual-report discrimination and confidence judgement report task (Figure 1D,E). This allowed us to test our new predictions for how confidence judgements should depend on the stimulus statistics at the same time as trying to replicate our earlier findings on changes in temporal biases. Ten frames of filtered noise with orientation energy at ± 45 degrees were presented to a participant, after which participants reported the dominant orientation.

3.2 Participants

Ten students at the University of Rochester (all naive to the goals to the study) participated in this study. All participants were compensated for their time. All experiments were performed by following the guidelines and methods approved by the Research participants Review Board.

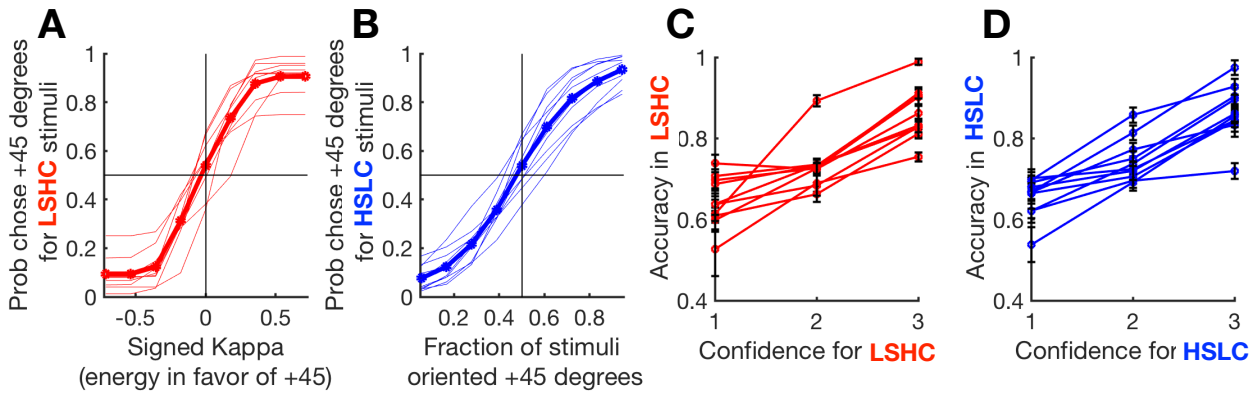

Figure 4: Relationship between accuracy and confidence reports. (A) Psychometric curves for LSHC condition. Thin lines denote individual participants and thick line indicates their mean. (x-axis) Positive valued κ denotes energy in favor of +45 degrees; negative values for -45. (B) Psychometric curves for HSLC condition. (x-axis) 0 corresponds to all stimulus frames consistent with -45 degrees; 1.0 means all stimulus frames consistent with +45 degrees. (C-D) Individual participant accuracy as a function of reported confidence for LSHC and HSLC conditions, respectively. All error bars are 1 std dev.

3.3 Stimulus

The stimulus was a band-pass filtered noise masked by a soft-edged annulus (Beaudot and Mullen, 2006; Nienborg and Cumming, 2014; Bondy et al., 2018). The annulus contained a small white cross in the center, on which participants were instructed to fixate. Each stimulus subtended 2.08 degrees of visual angle around the fixation point (see Figure 1D,E). The mean spatial frequency was $= 6.90$ cycles per degree, the spread of spatial frequency was $= 3.45$ cycles per degree, the (inverse) spread of orientation energy denoted by κ ranges from 0.0 to 0.8 (sensory information), the image luminance was 127 ± 22 and the width of the central annulus cutout was 0.43°. The number of frames that matched the correct answer for a trial, p_{match}, ranged from 0.5 to 1.0 (category information). We generated the stimuli using Matlab and Psychtoolbox and presented them on a 1920x1080px 120 Hz monitor with gamma-corrected luminance (Brainard, 1997). Participants kept a constant viewing distance of 105 cm using a chin-rest.

The design of the stimulus minimized the effects of small fixational eye movements. The range of spatial frequencies was kept constant for all participants (same as in (Lange et al., 2020)).

3.4 Procedure

In the LSHC condition, following (Lange et al., 2020), we run a 2:1 staircase on κ (starting from 0.8) which controls the amount of orientation energy in each frame of stimulus, and hence how hard to detect. We keep p_{match} fixed to 0.9 across all trials. Similarly for the HSLC condition, we run a 2:1 staircase on p_{match} (starting from 0.9) while keeping κ fixed to 0.8 such that the orientation in each frame of stimulus is clearly visible to participants.

Each trial starts by presenting a white cross in the center of a gray screen and a black circle outline for 200 ms, indicating the location where a series of stimuli will appear. Then 10 stimulus frames are presented around the cross, each lasting for 83 ms (12 frames per second). After the 10th frame, a noise mask with no orientation information is presented to prevent any after-images. Participants make a decision by pressing a button, indicating whether the majority of stimulus frames were oriented +45 or -45 degrees, and then report their confidence by pressing 1, 2, or 3 (1 is least confident, 3 is most) within the next 1 sec. Finally, auditory feedback was played at the end of each trial (see Figure 1D,E).

Participants learned the task using 50 trials of each condition. After that participants completed between 1500 to 2100 trials in the LSHC condition and 1000 to 1500 trials in HSLC condition. Trials were run in blocks of 100 trials,

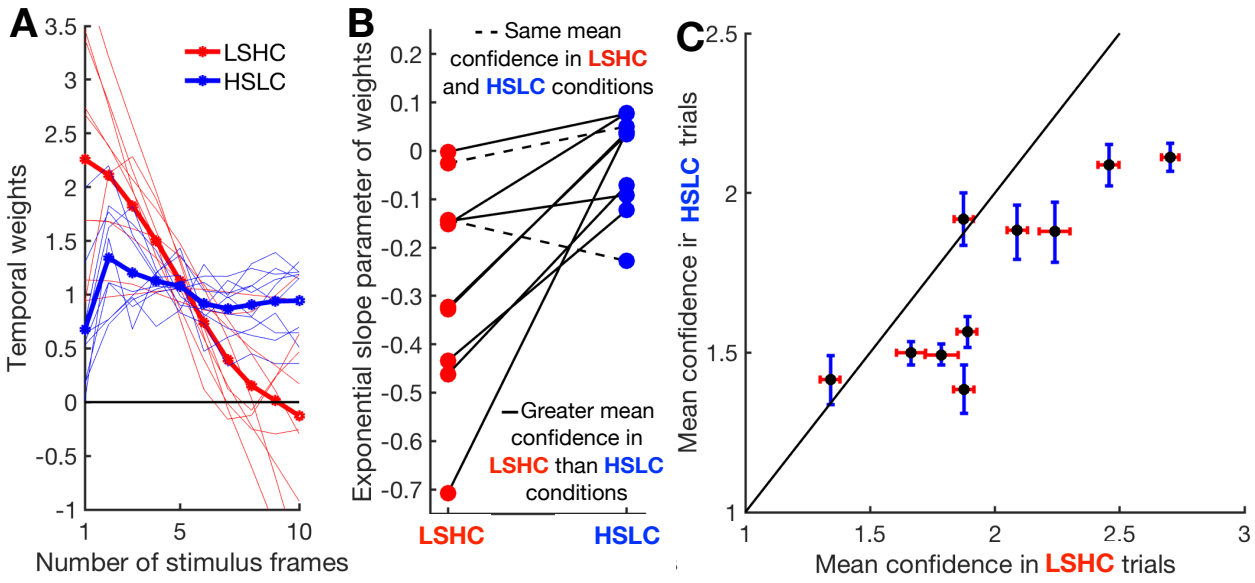

Figure 5: Participants showed higher confidence in LSHC task compared to HSLC task (A) Temporal weights showing primacy effect for LSHC condition (red) and optimal weighing in HSLC condition (blue). (B) Slopes of the weights change across the two tasks for each participant. Note that participants whose mean confidence reports lie on identity line in (C), dashed lines, show little or no change in slope. (C) Mean confidence reported for LSHC condition and HSLC condition. Most points lie below the identity line, indicating greater confidence in LSHC, confirming model predictions. Green dot shows participant mean and the error bars are 1 std dev. Performance is matched for comparison of LSHC and HSLC conditions (70%).

after which participants could take a small break.

3.5 Analysis

We calculate psychometric curves with respect to signal κ for LSHC and p_{match} for HSLC. We take the dot product of the Fourier-domain energy of each stimulus with a difference of Fourier-domain kernels at $\pm 45^\circ$ to compute signed signal κ. The signal is a scalar value that is positive if the stimulus contains more $+45^\circ$ energy and negative if it contains more -45° energy. Accuracy is determined by simply taking the average for all trials in which participants reported a given confidence values (Figure4).

Temporal weights of frames are computed for trials below the threshold performance (70%) for each participant. We fit the weights using regularized logistic regression and then normalized them to have a mean of one. We also compute a slope for the weights by fitting an exponential curve (Figure 5).

3.6 Findings

First, we replicated our previous findings (Lange et al., 2020): we confirmed that participants were able to perform the task as shown by psychometric curves (Figure 4A,B) and that the temporal weight profiles indeed showed a primacy effect for the LSHC condition and a recency effect for the HSLC condition (Figure 5A). We compared the change in temporal bias for each participant by comparing the slope of their weighing profiles (Figure 5B) and confirmed that most participants showed stronger recency in the HSLC condition.

We next confirmed the validity of the participants' reported confidence for both conditions (Figure 4C,D), by showing their that reported confidence, on average, tracked their accuracy.

Finally, we compared participants' confidence reports between the two conditions using trials under threshold performance (70%) (see Figure 5). We found that all but two participants were more confident in the LSHC condition, confirming our model predictions. Interestingly, the two participants who showed similar confidence in each trial (points overlapping the identity line) also showed small or negative effects with respect to changes in their temporal weighing slopes (dashed lines in Figure 5). This further supports the suggested close link between biases in confidence and temporal weighting of evidence in humans.

4 Discussion

Replicating and building on our previous work (Lange et al., 2020), we found that approximate hierarchical inference induces biases in explicit confidence judgements that are related to temporal weighting biases in predictable ways. Model simulations and new experimental evidence show that overconfidence co-occurs with the primacy effect when sensory information is low and category information is high.

This work suggests that approximate hierarchical inference may provide a computational basis for the emergence of biases beyond low-level perceptual decision-making. This model explains how biases in confidence, a higher-level cognitive faculty, can emerge due to a positive feedback loop between decision-making and sensory representations that depend on stimulus statistics in a systematic way.

Interestingly, the effect on confidence we found here differs with previous studies which examined the effect of stimulus volatility on confidence, showing that volatility increases confidence (Zylberberg et al., 2016; Castañón et al., 2019). While it may seem that the "volatility" in our stimulus corresponds to lower confidence, we hold that these experiments do not translate directly – it is an open questions as to how sensory and category information relate to stimulus volatility and what their combined effect on confidence is.

Finally, we speculate that the described dynamics and biases in confidence generalize from perceptual decision-making to the cognitive domain, with the key elements being temporal consistency of the evidence (category information) and the approximate nature of the brain's inference algorithms (Griffiths et al., 2015).

Acknowledgments

This work was supported by NEI/NIH awards R01 EY028811-01 (RMH, AC) and a Discover Grant for Undergraduate Summer Research from the University of Rochester (MS).

References

William HA Beaudot and Kathy T Mullen. Orientation discrimination in human vision: Psychophysics and modeling. *Vision research*, 46(1-2):26–46, 2006.

Adrian G Bondy, Ralf M Haefner, and Bruce G Cumming. Feedback determines the structure of correlated variability in primary visual cortex. *Nature neuroscience*, 21(4):598–606, 2018.

David H Brainard. The psychophysics toolbox. *Spatial vision*, 10(4):433–436, 1997.

Bingni W Brunton, Matthew M Botvinick, and Carlos D Brody. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.

Santiago Herce Castañón, Rani Moran, Jacqueline Ding, Tobias Egner, Dan Bang, and Christopher Summerfield. Human noise blindness drives suboptimal cognitive inference. *Nature communications*, 10(1):1–11, 2019.

Jan Drugowitsch, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, 92(6):1398–1411, 2016.

Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annual review of neuroscience*, 30, 2007.

Thomas L Griffiths, Falk Lieder, and Noah D Goodman. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science*, 7(2):217–229, 2015.

Piercesare Grimaldi, Hakwan Lau, and Michele A. Basso. There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, 55:88–97, 2015.

Roozbeh Kiani, Timothy D Hanks, and Michael N Shadlen. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12):3017–3029, 2008.

RD Lange, A Chattoraj, J Beck, J Yates, and R Haefner. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. biorxiv, 440321. 2020.

Hsin-Hung Li and Wei J. Ma. Confidence reports in decision-making with multiple alternatives violate the bayesian confidence hypothesis. *Nature communications*, 11(1):2004–2004, 2020.

Florent Meyniel, Mariano Sigman, and Zachary F Mainen. Confidence as bayesian probability: From neural origins to behavior. *Neuron (Cambridge, Mass.)*, 88(1):78–92, 2015.

Hendrikje Nienborg and Bruce G Cumming. Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 459(7243):89–92, 2009.

Hendrikje Nienborg and Bruce G Cumming. Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *Journal of Neuroscience*, 34(10):3579–3585, 2014.

Alexandre Pouget, Jan Drugowitsch, and Adam Kepecs. Confidence and certainty: distinct probabilistic quantities for different goals. *Nature neuroscience*, 19(3):366, 2016.

Valentin Wyart, Vincent De Gardelle, Jacqueline Scholl, and Christopher Summerfield. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron*, 76(4):847–858, 2012.

Ariel Zylberberg, Christopher R Fetsch, and Michael N Shadlen. The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *Elife*, 5:e17688, 2016.