Simulated voxels from the tuned inhibition model of perceptual metacognition to drive model validation via fMRI

Shaida P. Abachi (sabachi@uci.edu), Brian Maniscalco (bmanisca@uci.edu),
& **Megan A. K. Peters (megan.peters@uci.edu)**
Department of Cognitive Sciences, University of California Irvine
Irvine, CA, 92697, USA

Abstract

When we make perceptual decisions, confidence usually co-varies with decisional accuracy. However, sometimes this correspondence breaks down, e.g. in atypical environments or clinical populations. This raises an important question: what are the neural computations of perceptual metacognition if their output can diverge from perceptual decisions themselves? In a recent paper, we argued that tuned inhibition (TI)—i.e., the degree to which a neuron is inhibited by neighboring neurons with opposing tuning preferences, which varies from neuron to neuron—is a crucial part of the underlying mechanism. Here we explore how we might validate the TI model using fMRI data, by simulating the activity of 'voxels' of different compositions in the presence of evidence for and against a perceptual decision in a decision+confidence task. We show that we can quantify how a voxel's TI level dictates its predictive power for confidence judgments, providing support for use of these stimuli and analyses in fMRI data to validate the TI model of perceptual metacognition.

Keywords: perceptual decision making; perceptual confidence; metacognition; fMRI

Introduction

Our sense of confidence *usually* tracks our decisions' accuracy. However, disruption of this decision-confidence pairing has been linked to psychiatric disorders (Heinz et al., 2019), e.g. high confidence in wrong perceptual inferences may be responsible for hallucinations. We can also create decision-confidence dissociations in laboratory-based experiments, revealing that confidence computations track decision-congruent evidence magnitude and ignore evidence for the competing decision, i.e. a confirmation bias (Maniscalco, Peters, & Lau, 2016; Zylberberg, Barttfeld, & Sigman, 2012).

This confirmation bias in confidence may be due to the role of tuned inhibition (TI) in decisions and confidence (Maniscalco et al., 2021), wherein each sensory neuron's activity is differentially affected by surrounding network activity. In this model, units that independently reflect accumulated evidence of the preferred stimulus contribute to confidence judgments and have low levels of "inhibition tuning" (not inhibited by neurons with opposing preferences; Fig. 1A). In contrast, neurons that calculate relative evidence for a stimulus weighed against an opposing possibility contribute to decision, and have high levels of "inhibition tuning".

Here, we aimed to develop a way to identify inhibition tuning at the voxel level, so that this model could be critically tested in humans using fMRI. Thus, we built 'voxels' and simulated their activity in the presence of evidence *for* (*positive evidence*, PE) and *against* (*negative evidence*, NE) the correct stimulus alternative in a two-alternative forced choice task, resembling the stimuli we present to humans in the MRI scanner. These simulations revealed how we can identify voxels of different inhibition levels and quantify how they contribute to confidence computations using functional MRI data.

Methods

We produced quantitative predictions for voxels using simulations of the model described in (Maniscalco et al., 2021).

First, we designed random dot kinematogram stimuli that vary motion energy and conflict (Fig. 1B) in order to differentially activate x- and δ-neurons (these are also shown to human participants in the MRI scanner; data not shown). (We assume that, on average, a voxel contains equal proportions of left- and right-preferring neurons, so will be activated approximately equally for left or right motion.) δ-neurons and x-neurons were simulated from the model (including fitted parameters) reported in Maniscalco et al. (Maniscalco et al., 2021) with 100 trials per condition.

These 'neurons" activities were averaged across the simulated time of each trial and then combined to produce "BOLD"-like response signal R in each 'voxel' in conflict condition C according to:

$$R_C = w_x x + w_\delta \delta + w_k k \quad (1)$$

where w is the weight put on each type of neuron i ($\sum_i w_i = 1$), and x, δ, and k are the mean activity of x-, δ-, and inhibitory interneurons, respectively, over the course of the trial. We simulated 'voxels' with 0, 20, and 40% inhibitory interneurons, with the remainder of 'neurons' in each voxel made up of δ-neurons and x-neurons in proportions of 0:100% to 100:0% in steps of percents of 10 (see Fig. 1C for two such 'voxels'). Qualitatively, we should expect voxels primarily made up of δ-neurons ("δ-dominant voxels") to reduce their activity a lot in the presence of conflict due to strong inhibition between units with opposing preferences, while voxels primarily made up of x-neurons ("x-dominant voxels") should not (Fig. 1D).

From the simulated BOLD responses, we calculated a voxel's level of inhibition tuning (VIT); this will be used in fMRI data to capture an actual voxel's VIT based on its response R to the four conditions (Fig. 1B). We define VIT for each voxel in each energy condition E as:

$$VIT_E = \frac{R_{LC} - R_{HC}}{R_{LC}} \quad (2)$$

where, as above, LC is low conflict and HC is high conflict.

Finally, we simulated a "left versus right" + confidence behavioral task by extending the code from Maniscalco et al. (Maniscalco et al., 2021). We simulated x- and δ-dominant neurons' activities across 1000 trials in response to 4 stimulus combinations with PE:NE ratios of 6:1 (easy) and 3:1 (hard) (crossed with high and low energy), where the observer had to make decisions about the primary direction of dot-motion and then rate confidence. Confidence was computed according to the primary model (x-neurons drive confidence), and we also simulated an alternative control model in which δ-neurons drive confidence. We used area under the receiver operator characteristic curve (AUROC) to quantify how predictive a voxel's response R was of trial-by-trial confidence judgments under these two models, and also to quantify whether a voxel could predict left vs. right decisions as a control.

All simulations were performed in Matlab (Version 2021b).

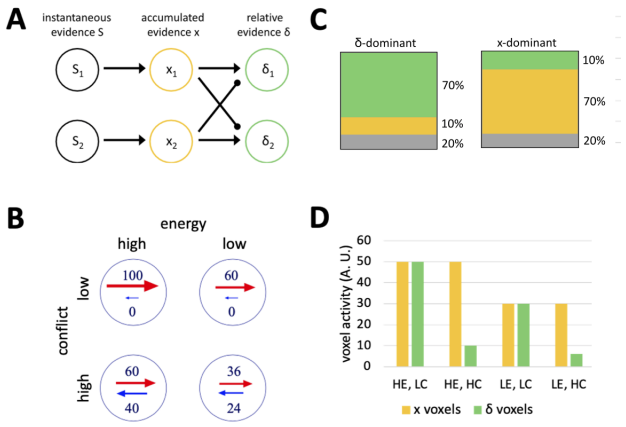

Figure 1: TI model, stimulus conditions, voxel composition, and qualitative predictions. (A) In the model, instantaneous evidence drives absolute evidence (yellow x-neurons), which excite units with similar tuning preferences but inhibit those with opposing preferences (green δ-neurons). Confidence is read out from decision-congruent x-neurons' activity. (B) Random dot motion conditions crossing low and high energy (LE & HE) with low and high conflict (LC & HC), designed to differentially activate x- and δ-neurons. (C) Sample simulated voxels with varying percentages of δ-neurons, x-neurons, and inhibitory interneurons. (D) Qualitative pattern of responding predicted from averaged activity of x- versus δ-neurons.

Results & Discussion

Voxels' responses R to low- and high-conflict stimuli show that the qualitative predictions are borne out quantitatively (Fig. 2A): while both x-dominant (less inhibited, low VIT, yellow) and δ-dominant (more inhibited, high VIT, green) voxels have higher activity under high energy conditions, only δ-dominant voxels exhibit strong reduction in R under high conflict relative to low conflict. Likewise, a voxel's percentage of x- or δ-neurons is well captured by VIT (Fig. 2B, Eq. 2). These results confirm the use of VIT as a scalar metric to capture a voxel's level of inhibition tuning based only on its pattern of responses to the four conditions from Fig. 1B.

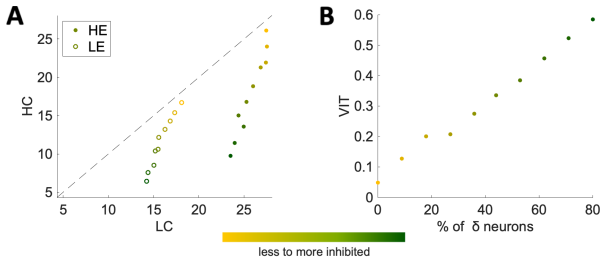

Figure 2: Simulation results. (A) Voxels' response patterns R depend on their makeup: x-dominant voxels do not change their activity under conflict, while δ-dominant voxels reduce their activity. (B) The x- versus δ-dominance of a voxel is captured by its *voxel inhibition tuning* (VIT) level (Eq. 2.)

The overall choice behavior of the simulated model data yielded perceptual decisions that were correct 66.3% of the time, with percent correct ∼70% for 6:1 PE:NE (easy) conditions and ∼62% for 3:1 PE:NE (hard) conditions. Crucially, we found that the primary (x-drives-confidence) model (Fig. 3, left panel) predicts that x-dominant voxels (low VIT) have higher predictive power for confidence judgments than δ-dominant voxels (high VIT). Neither the control model nor the left/right decision control shows this pattern.

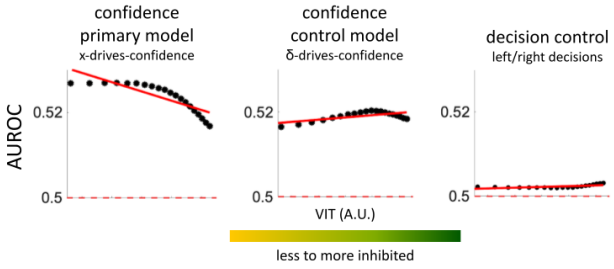

Figure 3: Voxels' predictive power for confidence depends on model specification. (left) In the primary (x-drives-confidence) model, VIT is inversely related to AUROC for confidence. (middle) In the alternative (δ-drives-confidence) model, VIT is unrelated to confidence AUROC. (right) VIT should never be related to predictive capacity for left/right decisions.

Our simulations demonstrate the validity of using the four dot-motion conditions in Fig. 1B to identify VIT at the voxel level, and then using AUROC to quantify predictive power for confidence as a function of VIT to validate the TI model (Maniscalco et al., 2021)—all using fMRI in awake, behaving humans. Ongoing work presents these four conditions plus a behavioral (choice + confidence) task to human observers in an MRI scanner while whole-brain BOLD signal is recorded, and will apply these analyses to validate the TI model (Maniscalco et al., 2021).

Acknowledgements

This work is supported by the Air Force Office of Scientific Research (award number FA9550-20-1-0106 to MAKP).

References

Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., & Waltz, J. A. (2019, September). Towards a unifying cognitive, neurophysiological, and computational neuroscience account of schizophrenia. *Schizophr. Bull.*, *45*(5), 1092–1100.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. K. (2021, March). Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLoS Comput. Biol.*, *17*(3), e1008779.

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016, April). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten. Percept. Psychophys.*, *78*(3), 923–937.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.*, *6*.