

Approximate inference explains paradoxical data in two-event causal inference task

Sabyasachi Shivkumar^{1,*}, Madeline Cappelloni²

Ross Maddox^{2,3}, Ralf M. Haefner^{1,*}

¹ Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA

² Biomedical Engineering, University of Rochester, Rochester, NY 14627, USA

³ Neuroscience, University of Rochester, Rochester, NY 14627, USA

* {sshivkum,rhaefne2}@ur.rochester.edu

Summary

The brain combines noisy and incomplete signals from multiple sources according to their reliability to infer the state of the outside world. The brain's implementation of this process of "probabilistic inference" is necessarily approximate. Here, we present theoretical insights and experimental results from a new causal integration task involving two auditory and two visual cues. While the auditory cues contain information about the correct choice, the visual cues do not. Despite the fact that the performance of an ideal observer does not depend on the location of the visual cues, human subjects' performance does. We show that this improvement can be explained by a model based on approximate inference (in our case sampling). Furthermore, we are able to quantify the "accuracy" of a subject's approximation using psychophysical data, something that is hard in simpler tasks in which sensor noise and inference noise affect behavior similarly. More generally, our task and model allow us to dissociate the three principal sources of sub-optimality in perceptual decision-making tasks: sensor noise (e.g. in photo receptors), model mismatch (mistaken assumptions about the structure of the world), and approximate inference. Depending on this partitioning, our model makes subject-specific predictions for how behavioral performance should scale with stimulus duration.

Keywords: causal integration, approximate inference, sampling

Introduction

The brain combines inputs from multiple senses to infer information about events in the outside world. The fact that it appears to do so flexibly, depending on the uncertainty associated with the respective sensory streams, is a central argument in favor of the 'Bayesian brain' hypothesis (Ernst & Banks 2002, Pouget et al. 2013). However, before it can combine different inputs, the brain has to decide whether two signals were actually due to a single event in the outside world or not – a process modeled as 'causal inference' (Kording et al. 2008, Acerbi & Ma 2017). Under natural conditions, the brain has to perform causal inference involving multiple cues associated with many potential causes. While previous studies concentrated on only one or two potential cause, our study extends

this work to multiple causes and cues – both experimentally and theoretically.

Task (Figure 1B)

Two brief (200ms) auditory stimuli, a pure tone and noise, are presented at equal eccentricity on opposite sides of the midline. The subject's task consist in reporting the side on which the tone appeared. Temporally paired with the auditory stimuli, random visual shapes are presented on the screen. The central comparison in our work is between two conditions called "matched" and "central". In the central condition the visual shapes are presented on the midline, in the center of the screen. In the matched condition, the visual shapes are presented at the same locations as the auditory signals. Importantly, the visual cues are random and *not* paired in any way with tone and noise and hence contain *no* information about the correct choice (left/right). (While the visual cues do contain information about the eccentricity at which the auditory tones are presented, the ideal observer does not benefit from them as shown below.)

Model (Figure 1A)

Generative model: We extended the single event causal inference model in Koerding et al. (2008) to the multiple cues and possible bindings in our task. The model assumes Gaussian noise where the subjects infer that the observation X_i could be generated from a location S_i with a variance σ_i where $i \in a - \text{tone}, a - \text{noise}, v - \text{left}, v - \text{right}$ where $a - \text{tone}, a - \text{noise}, v - \text{left}, v - \text{right}$ stand for audio tone, audio noise, visual cue on the left and visual cue on the right respectively. ϵ_a and ϵ_v correspond to the experimenter-chosen locations of tone and visual stimuli. $C \in \{0, 1\}$ represents the brain's belief about the correct causal structure, and $R \in \{0, 1\}$ the correct choice.

Exact inference: From the brain's perspective, performing the task here means computing the posterior over the location given its noisy sensory observations and reporting whether $p(R = \text{left}|\dots)$ was larger or smaller than $p(R = \text{right}|\dots)$. First, we derived an analytical expression for the posteriors over tone location, and over the correct choice (not shown here). We found that while this posterior depended on the location of the visual cues, this dependence never changed the optimal choice from the subject's perspective. Whenever

the posterior over tone location had more mass on the left of the midline than on the right of the midline, and whenever the posterior over correct choice being left was larger than the posterior over the correct choice being right, it stayed that way no matter the location of the visual cues. In other words, the behavior of an ideal observer who follows the optimal strategy of choosing the side with more posterior mass is unaffected by the visual cues, and hence the same for the central and the matched condition. What does change with visual cue location, however, was *how* much more mass was on the left compared to on the right (or vice versa). Figure 1C illustrates this finding by plotting the logarithm of the ratio of the posterior mass on the correct side compared to the incorrect side (aka the “log odds”) for one particular eccentricity comparing central and matched condition.

Approximate inference: Due to the intractable nature of the inference problems that the brain is faced with under natural conditions, its inferences are necessarily approximate. Deviations from exact inference (ideal observer) come in two principal class: **(1)** approximate representations and inference algorithms, and **(2)** model mismatch (Ma 2012).

(1) means that on any one trial, the brain only has access to an estimate of the posterior ratio based on its approximation to the posterior. As a result, performance will be decreased compared to an ideal observer who performs exact inference, with no “inference error” on its ratio estimate. (This does not mean that the ideal observer is perfect since it still bases its decision on noisy observations and hence its posterior is variable with more mass on the wrong side for some trials.) Furthermore, how likely this inference error will be in changing the ratio from above 1 to below 1 (the only case in which behavioral choice will be altered), will depend on the ratio itself. The further away the ratio from 1, the more likely the behavioral choice will correspond to the side with the larger mass. What we found in our model (and can show analytically for a simplified, forced-integration) model, is that whenever the noisy evidence implied a ratio > 1 for the central condition, it would imply an even larger ratio in the matched condition, and whenever it implied a ratio < 1 in the central condition, it would imply an even smaller ratio in the matched condition. This difference between the conditions has no effect for the ideal observer for which only bigger/smaller than 1 matters, but it has the effect that approximate inference performance is higher in the matched condition since the inference error is less likely to change the choice. This effect is illustrated in Figure 1C, and its presence in the experimental data is what allows us to constrain how close the approximation of the brain’s posterior is to the exact solution – independently from the amount of sensory noise with which it is usually confounded. One of the proposed approximation schemes that the brain may be employing is based on sampling (Fiser et al. 2010), and here we use the number of samples, n_s , as a measure for how accurate the brain’s posterior is. Importantly, our data collected

from 20 human subjects in this task allowed us to constrain n_s to relatively small values (most posterior mass over n_s given the data lies over 1-10 samples for most subjects – see Figure 1F for an example subject).

(2) means that the brain’s assumed model of the input signals, and correct choices deviates from the experimenter’s values. The largest deviation displayed by our subjects is a bias for one or the other choice. Interestingly, even in this case, performance is higher in the matched compared to the central condition. The intuitive reason for this is that integrating the visual cues with the auditory ones, increases the brain’s confidence about its sensory inputs (see case (1)), and hence decreases the relative influence of a bias in the prior belief about the correct choice. Importantly, this benefit of the visual cues is present even for the ideal observer, depending primarily on how much less noisy the visual information is compared to the auditory information, independent of n_s , helping to dissociate between those sources of errors (see Figure 1E).

Model fitting Using Gibbs-sampling in the generative model of the data, we have inferred the joint posterior over all the parameters in our model. The parameters in our model fall into three categories: (a) variabilities of the sensory inputs (noise), (b) prior parameters reflecting the statistics of the task, and (c) the accuracy of the posterior representation, n_s . The marginal distribution over any one parameter then yields our posterior belief about that parameter given the empirical data.

Subject-specific predictions: While our empirical data was recorded for only the two conditions described above, our model together with the inferred posterior over all parameters allows us to make subject-specific predictions for behavioral performance for any combination of visual and auditory eccentricity. Finally, at least under the assumption of a sampling-based representation, we can make subject-specific predictions on how the difference in performance between the central and the match condition should scale with stimulus duration: The inference noise should decrease with duration which presumably is proportional to the number of samples the brain can use (with an effect predicted by our model). In contrast, any improvement due to a subject bias should not scale in the same way as it is independent of the number of samples.

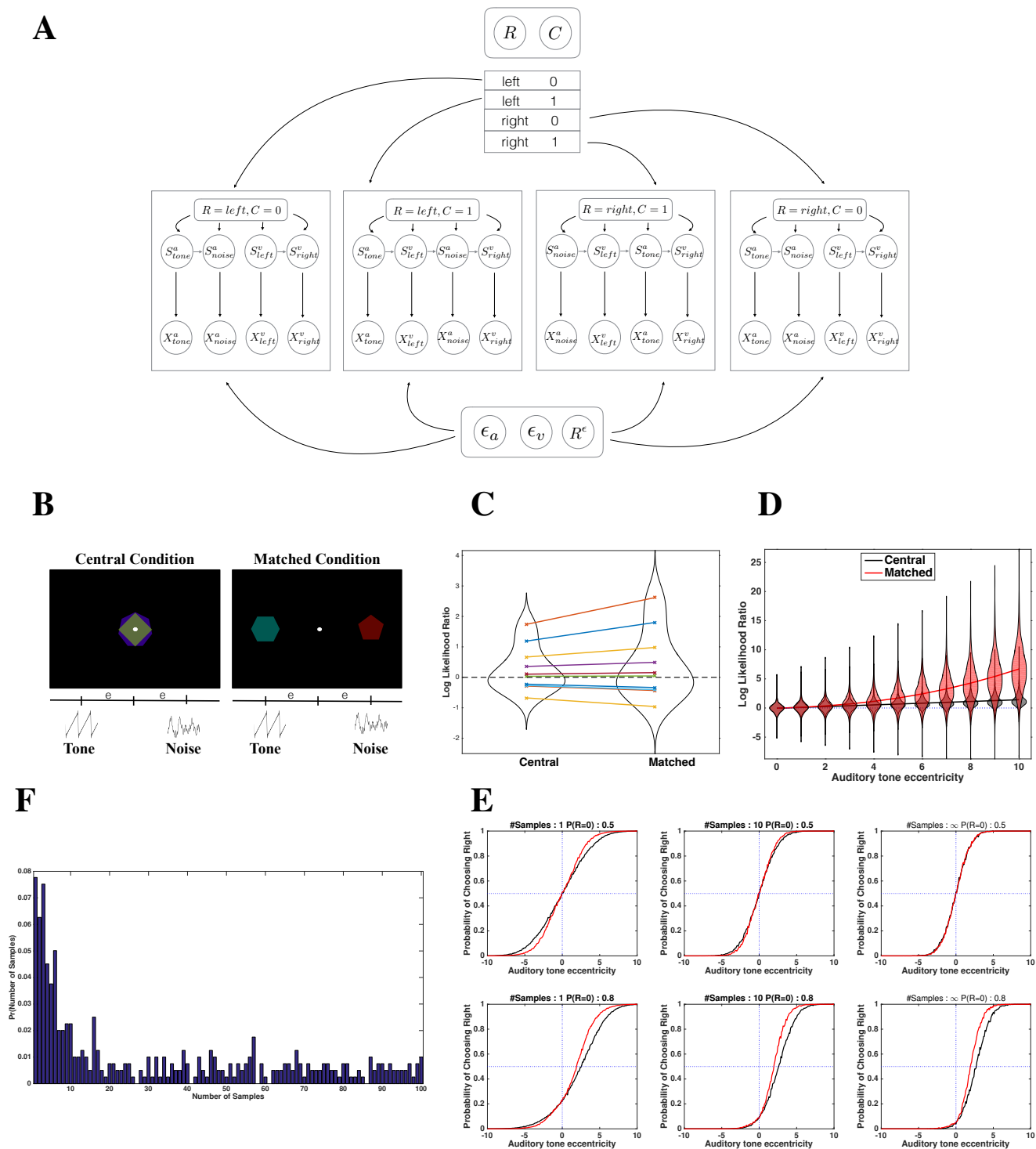


Figure 1: **A**) Generative model of the task. **B**) Illustration of display during task. **C**) Comparison of logarithm of posterior mass ratio (log odds) for central and matched conditions. Lines connect ratios for equal inputs, X . The log ratio always lies on the same side of 0 for both conditions. This means that the ideal observer would make the same choice in both conditions. Violin plots indicate the distribution of log odds due to variability (noise) in the actually observed signal on each trial. The relative mass in each violin below the x-axis is the same for both conditions. **D**) Log odds as a function of eccentricity. **E**) Psychometric curves in the absence (top row) and the presence (bottom row) of a choice bias. Left column corresponds to 1 sample ($n_s = 1$), 10 samples, and infinitely many samples (exact inference), respectively. Red curves represent matched condition, black curves central condition. **F**) Posterior over number of samples for an example subject. The posterior peaks for small values for n_s , clearly constraining the degree of accuracy of the brain's internal representation of its posterior over the auditory location/correct choice.