A representation-level algorithm for detecting spatial coincidences

Jennifer Lee (jenn.laura.lee@gmail.com)
New York University, 4 Washington Pl
New York, NY 10003, USA

Wei Ji Ma (weijima@nyu.edu)
New York University, 4 Washington Pl
New York, NY 10003, USA

Abstract

Spatial coincidences can lead to causal discoveries. We might expect to find a few ants on the sidewalk, but an unusually large cluster tips us off about the presence of a nearby food source. The leading cognitive explanation for this class of reasoning is Bayesian, but Bayesian models remain notoriously agnostic about the inner workings of the cognitive "black box." In this cluster detection paradigm, we ask what algorithms the brain might actually implement to detect spatial coincidences in an "approximately Bayesian" way. We find evidence that the brain represents two variables of the generative model: 1) the location of a hypothesized causal source and 2) the location of the points to which it gave rise. However, we propose that the brain is limited to representing probability distributions over one but not both of these variables, resulting in strong deviations from Bayes-optimality. We find, counterintuitively, that subjects become more prone to false alarms as the amount of information increases, and our proposed cognitive algorithm accounts for this pattern. Our representation-level algorithm elucidates the cognitive processes underlying coincidence detection, and helps explain our tendency to perceive causal patterns where none exist.

Keywords: Causal inference; Coincidence; Probabilistic reasoning; Perceptual grouping; Bayesian models

Introduction

Griffiths and Tenenbaum have proposed that a sense of coincidence can lead to causal discoveries (Griffiths & Tenenbaum, 2007)– for instance, in the "London bombing problem," an individual might look at a map of bombings to determine whether they are indiscriminate or targeted. Their normative Bayesian framework provides a highly unifying account of our sense of coincidence in a variety of contexts. But while the framework provides an approximate "as if" description of human behaviour, it falls short of making any commitments about the mental representations and algorithms carried out by the brain during this assessment. Indeed, if we were to translate the Bayesian model for spatial coincidence detection into a representation-level model of the inner workings of the brain, we expect that the number of computations required to solve a simple spatial coincidence detection task would quickly exceed a number which might plausibly be implemented by the brain.

In a decision-making task, a nuisance parameter is a variable that does not bear directly on the decision, but that must be accounted for in order to arrive at the variable of interest. Previous studies on category learning (Fleming, Maloney, & Daw, 2013) (Murphy, Chen, & Ross, 2012) and perceptual decision-making suggest that subjects might use simplified point-estimates of intermediate nuisance parameters instead of marginalizing over their full probability distributions (though see (Shen & Ma, 2016)), resulting in particular patterns of suboptimal behaviour.

In the current study, we employ a spatial coincidence task inspired by the "London bombing" problem. Our version of the task uses the spatial distribution of pigeons in a park, affected by a pigeon feeder whose location is not directly observable. Pigeons cluster around the pigeon feeder, if she is present. The subject's goal is to infer the presence or absence of the feeder. The generative model of the task entails two abstract parameters: 1) the location of the causal object (feeder) and 2) which of all observations "are affiliated with" (i.e., "result from") the causal object. We ask whether these two parameters are represented by the brain at all, and if so, whether they are represented in full probabilistic form, or as collapsed point estimates.

Our analyses rely on two important assumptions:

Firstly, we assume that practice trials and explicit verbal and graphical instruction are sufficient for subjects to learn the correct generative model. This involves, for instance, the assumption that subjects learn the general statistics of where and how often the bird feeder appears.

Secondly, we assume that the variables entailed by the generative model (e.g., the location of an unobserved object) must be represented by the brain as either a single-point estimate (e.g., the number '5'), or as a full probability distribution (e.g., a 'heatmap' of probability over multiple locations), or else not represented at all.

We then test three hypotheses about probabilistic representation in the brain during a spatial coincidence detection task. On the Strong Bayesian representation hypothesis, the brain actually represents all of the abstract parameters of the generative model, including their full probability distributions. On the Non-Probabilistic representation hypothesis, the abstract variables of the generative model are not mentally represented at all: instead, subjects assess spatial coincidences using some heuristic metric like the mean distance between points. Lastly, the Weak Bayesian representation hypothesis holds that the abstract variables entailed by the generative model are indeed represented by the brain, but that not all such parameters can be represented as full probability distributions: at least some are represented as single-point estimates.

In a variety of perceptual and cognitive tasks, much effort has been spent characterizing performance as "Bayesian" or "non-Bayesian" (taken broadly to mean any "suboptimal" heuristic strategy) without asking what it is the brain represents. The present study advance a new way of analyzing this vast category of suboptimal perceptual strategies by asking more specifically whether cognitive systems represent any of the abstract variables entailed by the generative model at all, and if so, in what form.

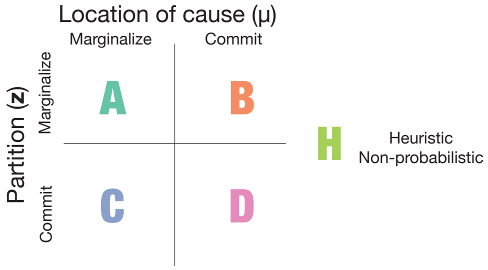

Figure 1: Taxonomy of model families. Family H models do not posit the representation of either location or partition.

Methods

Task 10 subjects were given a cluster detection task, in which they were asked whether a set of dots was drawn from a random uniform distribution, or from a mixture of a uniform distribution and a Gaussian. In task context, dots denoted the location of pigeons in a park. A "causal object" was introduced as an invisible "pigeon feeder" whose location was not directly observable. Subjects were instructed as follows: "On days when the pigeon feeder is present, pigeons tend to cluster around her location. But even when the feeder is present, there's only a 50% chance that a given pigeon will be affiliated with her." Pigeons which were not affiliated with the feeder were drawn randomly. The location of the feeder herself was drawn from a Gaussian distribution centered at the center of the screen. Subjects indicated whether a feeder was present by pressing a button. Their decision was based on the spatial distribution of pigeons on the screen. Subjects completed 20 practice trials with full feedback ("correct"/ "incorrect," including the partition of pigeons and actual location of the feeder if there was one), followed by the main task, comprised of 2000 trials with partial feedback ("correct"/ "incorrect").

The generative model In our task, two nuisance parameters are entailed by the generative model: μ, the location of the bird feeder, and $\mathbf{z}$, the "partition" of the pigeons. The partition denotes which of all pigeons are affiliated or unaffiliated with the feeder. It can be represented as a binary vector (affiliated/ unaffiliated) of length N (where N is the number of pigeons on that trial).

The ideal Bayesian observer would represent both μ and $\mathbf{z}$ as probability distributions. In the case of μ, this representation might be something like an imagined probability heatmap over possible locations where the feeder may be (e.g., an imagistic representation). In the case of $\mathbf{z}$, such a probability distribution is difficult to represent imagistically, as one would need to simultaneously represent the probability of every possible combination of independently drawn pigeons being affiliated or unaffiliated with the feeder (2^N possibilities). The brain may still represent it in some form nonetheless.

Observer models We test five families of observer models, each with a different assumption about what it is that cognitive systems represent during the task (Fig 1).

In Family A (the Strong Bayesian model), the brain represents full probability distributions over both variables of the generative model.

In Family B, the brain commits to and represents only a single partition ($\mathbf{z}$). That is, it only represents one of all possible combinations of affiliated pigeons. How exactly it chooses that particular set of "tagged" pigeons to represent as "affiliated" is determined by the specifics of each of the models within Family B. The subject still represents (and later marginalize over) the feeder location (μ) as a probability distribution over all possible locations.

Conversely, in Family C, the brain commits to and represents only a single feeder location (μ), but represents (and later marginalizes over) the full probability distribution of partitions ($\mathbf{z}$). For instance, in one model, the brain simply chooses the center of mass of all of the pigeons as its committed μ.

In Family D, the brain represents both feeder location and partition as point-estimates, and not as probability distributions. For instance, one model commits to both feeder location and partition by maximizing the joint posterior.

Lastly, in Family H, the brain does not represent feeder location or partition at all. Instead, subjects solve the task by representing some other abstract variable outside of the generative model, like "pigeon density." For instance, the subject might simply represent the density of the points on screen and respond "feeder present" if that number exceeds some threshold.

Family A tests the Strong Bayesian representation hypothesis, family B, C, and D test the Weak Bayesian representation hypothesis, and family H tests the non-probabilistic representation hypothesis. See Figure 5 for the full list of observer models by family.

Results

We first characterize behavioural responses as a function of basic stimulus properties like "pigeon density." As one might expect, subjects are more likely to respond "feeder present" the more densely packed the pigeons are on screen. Intuitively again, when the distance between the two closest pigeons is very small (i.e., when the two closest pigeons are

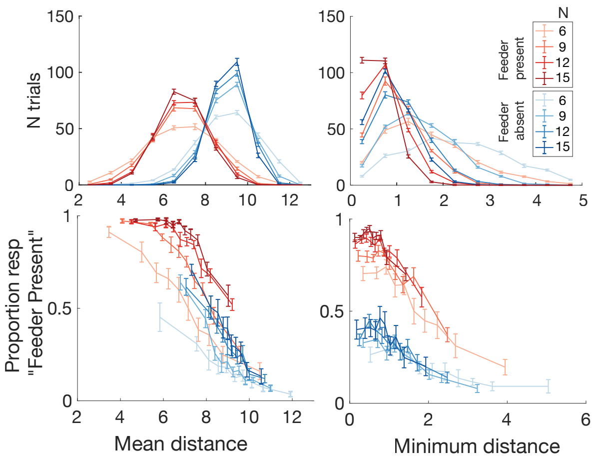

Figure 2: Top row: Histogram of stimulus distributions. Bottom row: Proportion of "feeder present" responses as a function of each distance-based heuristic. Colour saturation indicates the number of pigeons on a given trial.

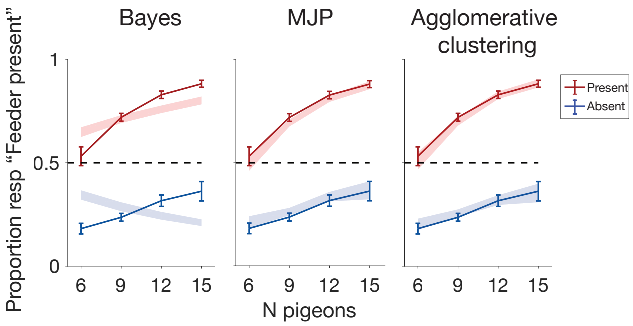

Figure 3: Model fits of proportion of "feeder present" responses as a function of number of pigeons (N) for three models: strong Bayesian, maximum joint posterior, and agglomerative clustering. The Bayesian model incorrectly predicts a decreasing effect of N on false alarms.

very close together), subjects are also more likely to respond "feeder present." (Fig 2) Nevertheless, subjects are surprisingly adept at telling apart trials when the feeder truly is or isn't present even when statistics like the "mean distance" between points are held constant. This fact is evidenced by the vertical offset of the red and blue curves in Fig 2. Therefore, from the outset, we see that if these abstract variables like density or minimum distance are represented, they cannot possibly be the only thing that subjects represent in order to make accurate judgements about spatial coincidence in this task.

One unexpected finding is that the proportion of "feeder present" responses increases as a simple function of the number of pigeons (N) on screen (Fig 3). In other words, as the number of uniformly drawn dots increases, subjects are more likely to sound a false alarm. Examining model predictions, we see that this unexpected qualitative trend is at odds with the Strong Bayesian hypothesis (Fig 3)– intuitively, if we were able to reason optimally, observing more dots drawn from a uniform distribution should make us more certain that the distribution is in fact uniform. Instead, we see that the number of "feeder present" responses increases with increasing N. We see this qualitative divergence from Bayes-optimal as the upward slope of the blue curve in Fig 3 against the downward slope of the curve predicted by the Bayesian model. Henceforth, we refer to the slope of this curve as the "effect of N on false alarms."

To summarize performance of the full range of models, we show two of the most informative aspects of model performance: the model's predicted effect of N on false alarms, and the overall predicted subject accuracy (% correct) in Fig 4. While the strong Bayesian model (Family A) predicts subject accuracy reasonably well, it is one of the only that predicts the wrong direction of the effect of N on false alarms. Family H model predictions are even farther afield.

Formal model comparison (Fig 5) shows that each of the three remaining "Weak Bayesian" families (B,C, and D) present at least one strong contender for winning model. From this, we are left only with variants of the Weak Bayesian hypothesis: the two abstract variables entailed by the generative model are indeed represented by the brain, but both cannot be represented as full probability distributions: at least one (if not both) are represented as single-point estimates.

Discussion

From our data, we can conclude that subjects likely do represent both variables of the generative model in a spatial coincidence detection task, but that at least one of the two cannot be represented as a full probability distribution.

Importantly, there is a reasonable way to distinguish between the top four Weak Bayesian models, even though they are functionally indistinguishable as per AIC and BIC comparison (that is, even though they each provide statistically indistinguishably good fits to the data). We propose analyzing the plausibility of each model at the representational level. Only one the four winning models evades the combinatorial explosion of unique representations and provides a plausible model for mental representation in the brain: Agglomerative Clustering.

In Agglomerative Clustering, the observer commits to a particular $\mathbf{z}$ rather than representing a full probability distribution. We choose this $\mathbf{z}$ by picking a single point as the cluster "seed." The cluster is hypothesized to belong to the causal source. One iteratively adds the next-nearest point to the cluster, each time evaluating the log likelihood ratio which results from that particular $\mathbf{z}$. We continue adding points until the log likelihood ratio no longer increases, resulting in some spatially contiguous set of points hypothesized to belong to the source if such a source exists, represented by the committed $\mathbf{z}$. The

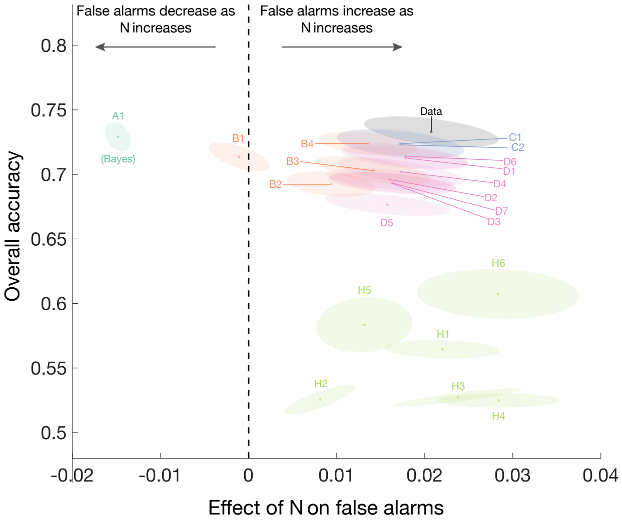

Figure 4: Two informative dimensions of model predictions: overall predicted subject accuracy, and effect of N on false alarms (the slope of the blue shaded line in Fig 3)

$\mathbf{z}$-dependent log likelihood ratio is then calculated as a decision variable with noisy threshold. Importantly, the number of operations required to find this $\mathbf{z}$ is less than N^2. Crucially, this algorithm coheres with intuitions that the least plausible possibilities of $\mathbf{z}$ are not represented or considered at all during the decision-making process.

While still speculative, the notion that such probabilistic inferences might be mediated solely by analog, iconic representations is attractive, and agglomerative clustering satisfies this constraint. A mental representation of the commitment to a particular $\mathbf{z}$ might simply entail the tagging or highlighting of certain points in a display. From that representation, all that remains is to marginalize over all possible locations of the feeder, which plausibly amounts to some operation on an imagistic heatmap-like representation. This magnitude would finally be compared against a noisy decision threshold.

The value of this reduction to an iconic representation-level algorithm is clear: while the Bayesian model provides a general computational lens for viewing the problem to be solved, it does not provide us with a plausible picture of the mental representations and operations which implement the computation. Furthermore, we find a qualitative divergence of the Bayesian prediction from human data– namely, increasing the number of observations increases the rate of false alarms. We make progress by pursuing a plausible representation-level algorithm which accounts for this effect, narrowing the search space of causal mechanical explanations for the phenomenon.

There are a number of limitations to this approach. Modelling behavioural evidence will never provide definitive proof that any one variable is necessarily represented by the brain. Variables can be represented without ever being used downstream to arrive at a decision. Nevertheless, the current study provides considerable evidence to favour the Weak Bayesian account of representation, and the further hypothesis that suboptimal judgements of spatial coincidence may be accounted for by natural limitations in iconic representation. Further evidence for the existence of particular iconic representations should come from paradigms which test representational "transfer" (discussed in (Maloney & Mamassian, 2009)): an abstract posit of a computational model is more likely to truly be represented in the brain if it is also shown to be used in a subsequent, disparate task.

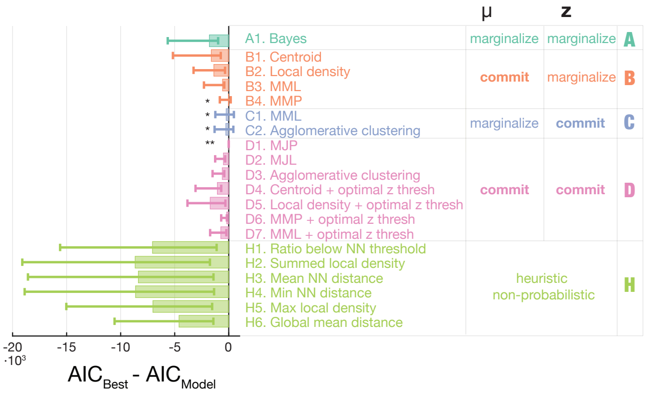

Figure 5: Model comparison. MML denotes "Maximum Marginal Likelihood," MMP denotes "Maximum Marginal Posterior," MJP denotes "Maximum Joint Posterior." NN denotes "Nearest Neighbour." The four winning models (starred) are indistinguishable via AIC and BIC.

Acknowledgments

Funding for this project was provided by NSERC (CGSD3-516810-2018).

References

Fleming, S. M., Maloney, L. T., & Daw, N. D. (2013). The irrationality of categorical perception. *Journal of Neuroscience*, *33*(49), 19060–19070.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition*, *103*(2), 180–226.

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing bayesian transfer. *Visual neuroscience*, *26*(1), 147–155.

Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning*, *18*(1), 81–117.

Shen, S., & Ma, W. J. (2016). A detailed comparison of optimality and simplicity in perceptual decision making. *Psychological review*, *123*(4), 452.