**Central contribution:** Monkeys and humans use similar strategies for multisensory causal inference.

**Key points of this paper:**

1. Important conceptual gaps between theoretical (single neuron), neural (whole brain fMRI and MEG), and behavioral level understanding of causal inference
2. We developed a task that can be used by both monkeys and humans
3. Humans behavior is consistent with previous descriptions of causal inference in similar tasks
   1. Unity
   2. Localization
   3. Joint
4. Monkey behavior is consistent with human behavior as well as XYZ models of causal inference
5. (maybe) demonstrate that enough behavioral data can be collected in a single session for ephys
6. Compare with other typical models for validation

Blue text is outline: 1 informal sentence per section, then 1 informal sentence per paragraph, then fill in.

Context-content-conclusion

**Abstract**

**Intro – We need monkeys to do this kind of task, so we can poke into their brains and see what’s going on**

Information from multiple senses can be combined to improve accuracy of sensory inference, but this is only useful if all combined sensory streams reflect information about the same source.

Perception is inherently multisensory, with multiple discrete sensory streams being processed and combined to produce a coherent representation of the world that most effectively guides behavior. This is particularly obvious in conditions where information from one sensory modality can be used directly to reduce uncertainty about another, such as reading the lips of a speaker to improve language comprehension. However, this integration of sensory modalities is only beneficial if the two sensory streams originate from the same source in the external environment.

Humans solve this problem in a Bayes optimal manner, which has been thoroughly explored in the literature.

This problem of determining whether two noisy observations originate from the same source is a form of causal inference (CI). Recent behavioral research has demonstrated that human observers perform this task in a manner consistent with an ideal Bayesian observer (Acerbi, Dokka, Angelaki, & Ma, 2018; Dokka, DeAngelis, & Angelaki, 2015; Dokka, Park, Jansen, DeAngelis, & Angelaki, 2019; Körding et al., 2007; Rohe et al., 2015; Wozny, Beierholm, & Shams, 2010). Broadly, causal inference models a type of hierarchical inference that compares the relative likelihood of the two causal scenarios (same source or different sources) for generating an observed pattern of sensory input in order to determine which causal scenario is most likely (figure 1A). The resulting inference can then be used to either to make a judgement about the number of discrete sources (explicit causal inference) or influence how the sensory input is perceived (implicit causal inference).

Understanding of this phenomenon at the level of individual neurons has been challenging, because to date only human subjects have been used to look at CI, and we can’t easily record single neuron activity from them.

There is a wide gulf between our understanding of causal inference at the behavioral level vs. at the neuronal level. Theoretical attempts to implement causal inference at the level of neural circuits have met with mixed success (Cuppini, Shams, Magosso, & Ursino, 2017; Fetsch, DeAngelis, & Angelaki, 2013; Ma & Rahmati, 2013). At the much higher level of cortical processing, it has been suggested that multisensory causal inference may be accomplished in a hierarchical fashion, with lower level sensory cortices representing unisensory information (segregated), other brain regions reflecting fused stimuli (integrated), and finally culminating in complete causal inference at the level of either pre-frontal cortex or intraparietal sulcus (Cao, Summerfield, Park, Giordano, & Kayser, 2019; Rohe & Noppeney, 2015, 2016). This view of hierarchical neural processing is pleasingly consistent with the hierarchical nature of ideal observer models of causal inference. However, it is inconsistent with other research that shows significant interaction between modalities even in primary sensory and subcortical regions [refs]. It’s possible that this conflict is due in part to the level of experimentation, with the previously discussed neuronal findings relying on human neuroimaging (fMRI and MEG) which cannot detect multisensory effects occurring at the level of single neurons. This question is most easily answered using animal models, but to date behavioral paradigms which produce measureable indications causal inference in non-human animals have been almost non-existent (though see Dokka, Park, Jansen, DeAngelis, & Angelaki, 2019).

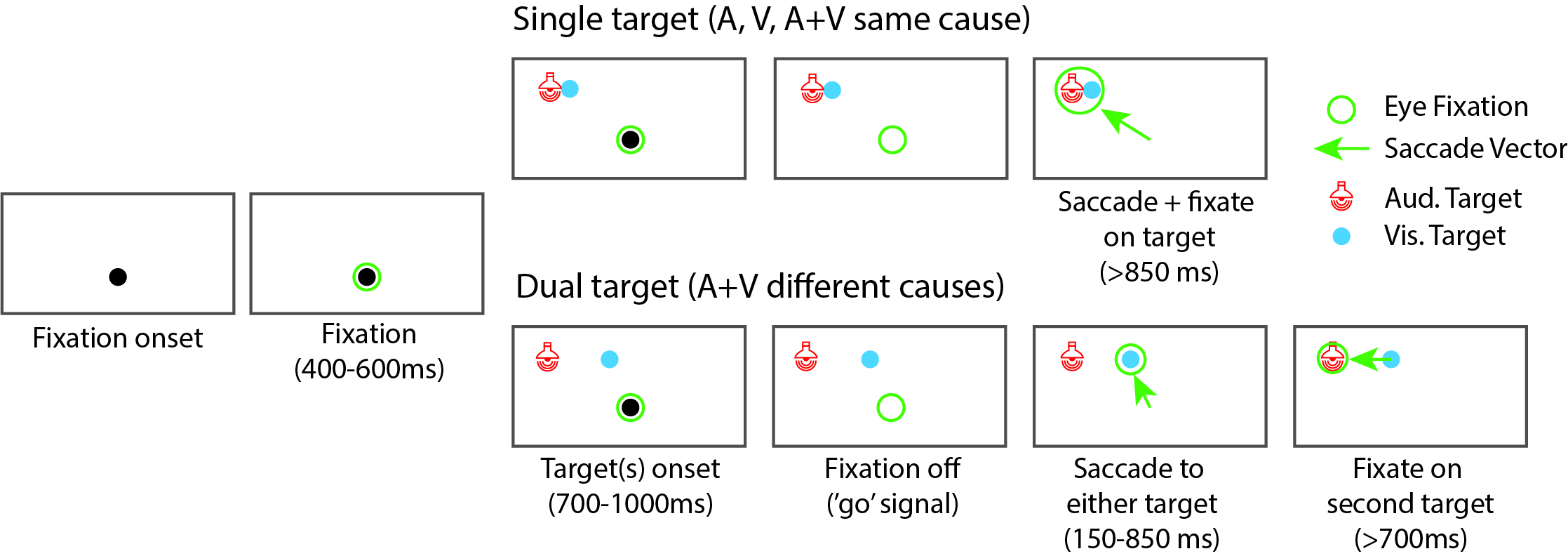
Here we create a behavioral paradigm that requires both implicit and explicit causal inference, which can be used in both monkeys and humans, and find that monkeys appear to apply the same type of Bayesian causal inference strategy seen in humans. (validating them as a potential model organism for the study of causal inference at the single neuron and neural circuit level)

We developed a new behavioral paradigm which requires both explicit and implicit causal inference, and can be used in both humans and non-human primates. We find that human subjects perform the task in a Bayes optimal manner, consistent with performance in other similar tasks. Additionally, we find that monkeys show comparable behavior to human subjects, and that their behavior reflects similarly optimal causal inference strategies, with some important differences. Specifically…

The dual nature of our task also allows collection of sufficient data within a single experimental session, critical for pairing this behavior with electrophysiological experiments in animals. Together these results suggest that non-human primates are a viable model organism for beginning to understand the neural basis of causal inference at the single cell level.

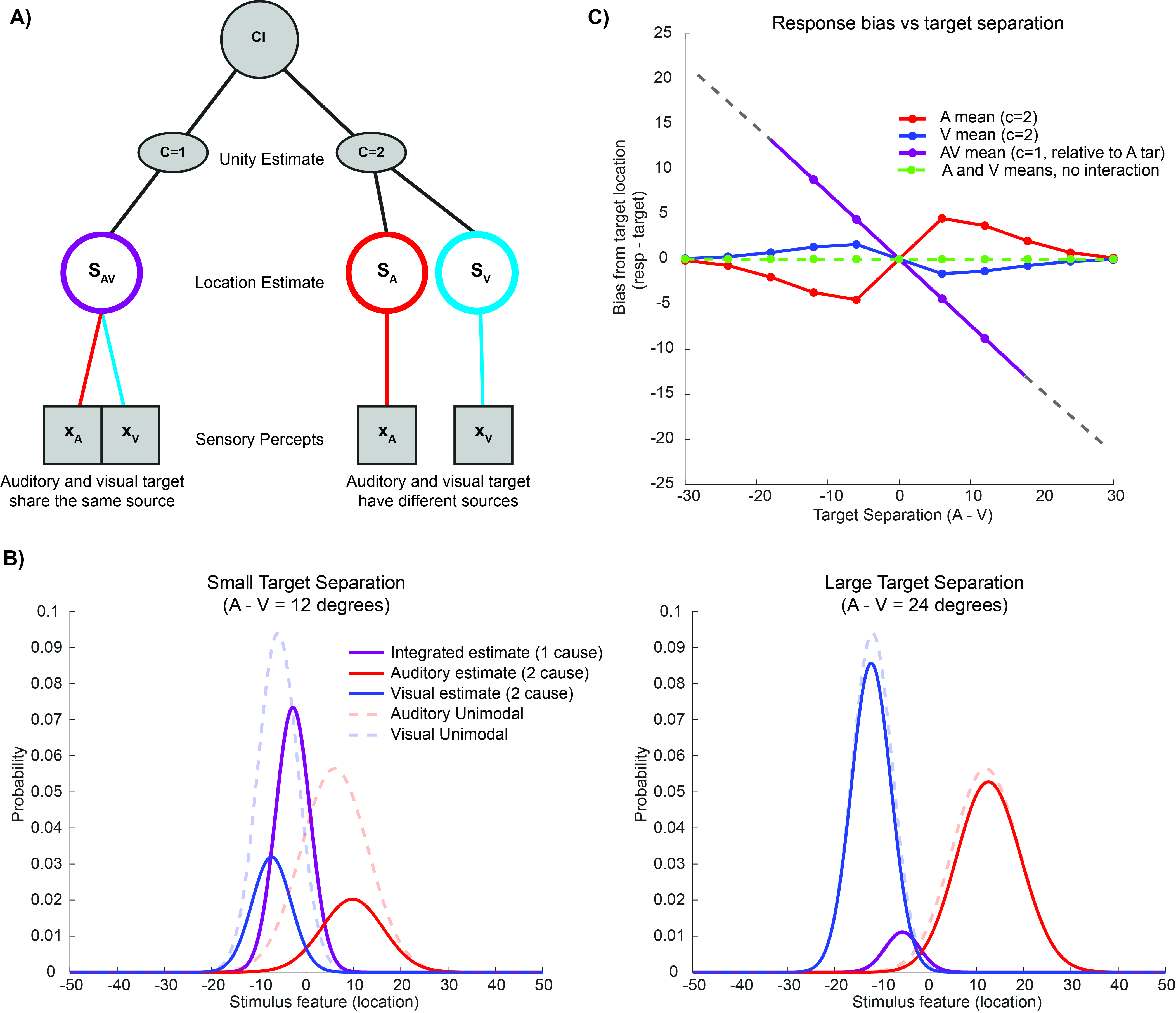
**Results – Monkeys can do it, and seem to be basically doing it like humans with some caveats**

We have created a dual causal inference task, one which requires that subjects report both the number and unique stimuli as well as their position, which relies on both auditory and visual stimulus information. (figure 1, behavioral paradigm description)



**Figure 1**: *Behavioral paradigm description.* Each trial begins when the subject holds fixation at a central light for a variable interval. This is followed by a variable stimulus presentation interval, where the subject is required to maintain fixation at the central location. After the go cue (extinguishing of fixation light), the subject indicates perception by making saccades to the sensory target(s). For single target trials (top: either unisensory trials, or trials with coincident auditory and visual stimuli), subjects make a single saccade to the perceived location and then hold fixation at that point until the end of the trial (at least 850ms). For multiple target trials, subjects make two saccades in rapid succession to each target in any order. These two trial types were interleaved throughout the recording session.

[framing sentence to be added]. We developed a novel multisensory causal inference paradigm, building on work from previous research (Körding et al., 2007; Rohe & Noppeney, 2016; Wozny et al., 2010). Subjects were seated in a dark, anechoic chamber facing a row of co-located speakers and LEDs. Trials were randomly interleaved and consisted of either unisensory (single auditory or visual stimulus), or multisensory (auditory and visual stimuli, played at the same time and for the same duration) stimuli. Multisensory trials had various amounts of spatial separation between the auditory and visual targets, ranging from 0 degrees (coincident) to 36 degrees. On every trial, subjects were asked to report the location of both the auditory and visual stimulus. For unisensory or multisensory-coincident trials (figure 1, top panels), subjects made a single saccade to the perceived location of the stimulus source and then held fixation at that point. For multisensory-separate trials, subjects made two saccades in rapid succession, one to each of the perceived sources. This dual task design allowed characterization of both explicit (one vs. two saccades) and implicit (location of fused percept) causal inference.



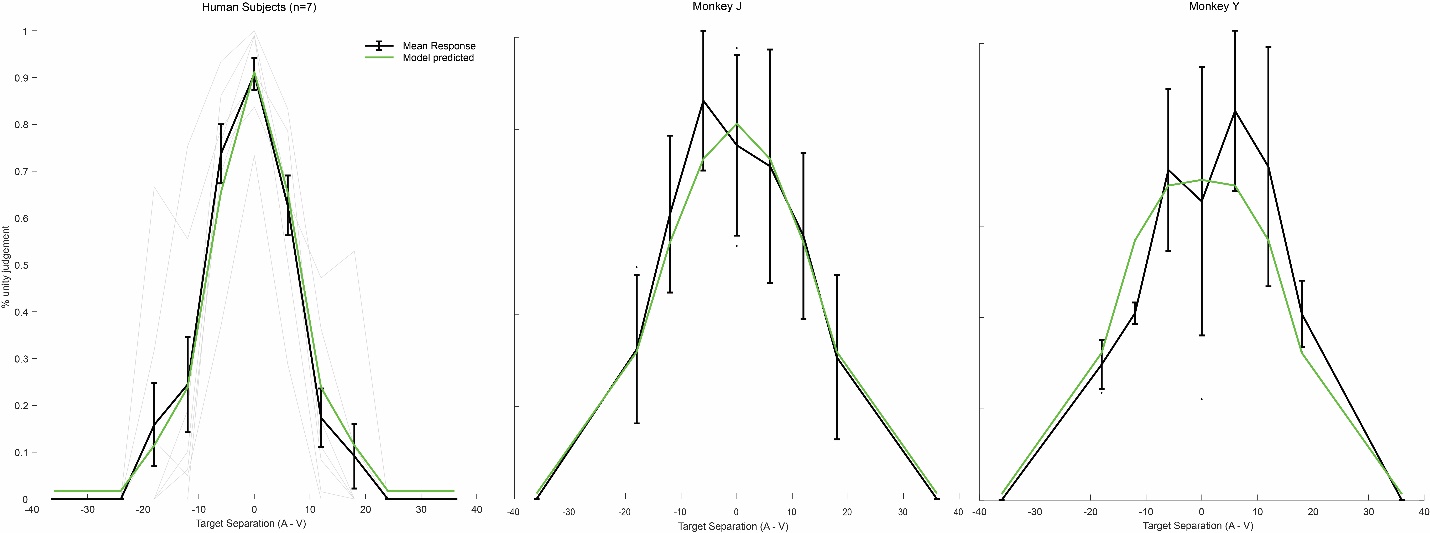
**Figure 2:** *Schematic descriptions of causal inference* **A)** Sensory percepts ( arise from stimuli in the world and are assumed to originate from either the same cause (though perturbed by different amounts of sensory noise) or from independent causes. Depending on the causal scenario (c=1, left branch, or c=2 right branch) these sensory percepts are used to produce location estimates (. The observer compares the relative likelihood of each of the potential causal scenarios in order to determine which was most likely to produce the observer sensory percepts, and may use this estimation in order to reweight or arbitrate between location estimates. **B)** Estimated probability distributions for the source of sensory percepts under the one cause (, purple) or two cause (, red or , blue) conditions. The distributions are normalized such that the total area under the solid curves sum to one, so that the relative ratio between one and two saccade trials can be seen by comparing the left and right panels. In addition, the forced causal decision on every trial results in a negative bias when compared to the estimates from unimodal trials (solid vs dashed curves) that depends on target separation (left vs right panels). **C)**  Under the causal inference model, single saccade responses show a bias (relative to the location of the auditory target) that linearly depends on target separation (purple solid line). The slope of this line depends on the ratio between auditory and visual sensory variance. Double saccade responses (red and blue curves) show a bias in the opposite direction, which depends both on the spatial separation between stimuli and on the sensory variance. Observers which do not perform causal inference do not display this separation dependent bias (green dashed line).

We have modeled this by adapting several forms of ideal observer models that have been previously applied to human behavioral experiments, including both optimal and heuristic decision rules. Because it is not strictly necessary that subjects use the same causal inference strategy for both localization and unity judgement, we combined models in a factorial manner. (fig 2, model descriptions and predictions)

We model both the explicit and implicit causal inference features of this task by adapting the models common in the literature (figure 1A) (Acerbi et al., 2018; Körding et al., 2007; Wozny et al., 2010). In this ideal observer model, it is assumed that some sensory stimulus in the world (in our case, a sound or a light) produces a noisy internal representation ( or , respectively) which is interpreted under one of two causal scenarios. In the single cause case, the two internal variables are interpreted as independent samples which are informative about the same object, and therefore can be integrated according to improve localization accuracy (Alais & Burr, 2004; Ernst & Banks, 2002). In the two cause case, however, the sources are considered to be unrelated and therefore there should be no interaction between the percepts. The observer can use the relative likelihood of each of these scenarios to either estimate whether the one cause or two cause case is most likely (c=1 or c=2, explicit causal inference), or to most accurately localize the source(s) of the stimuli (SAV if one cause, SA and SV if two causes, implicit causal inference). Importantly, both of these features must be estimated by subjects on every trial, resulting reported location distributions that can be compared directly with the estimated distributions under each of the causal structures.

Because of the dual nature of our tasks, subjective reports provide information about both the estimated location of the source(s) under each causal structure (figure 2B, solid curves) as well as the likelihood of perceived one source vs. two sources (ratio of one-saccade to two-saccade trials). This has several important consequences. First, the response distribution for small target separation values will consist mostly of single saccades, while the responses for large target separations will consist mostly of double saccades (figure 2B, compare area under purple curve for left vs. right panel). Second, the causal judgement will induce a bias in the mean of the two-saccade auditory and visual distributions compared to the same estimates on unimodal trials. This paradoxical negative bias is caused by the fact that the noisy internal representations which are most similar to one another (i.e., perceived closer together) are reported via a single saccade. This results in an effective selection bias that will shift the reported distributions in the opposite direction from what is expected under sensory fusion(Körding et al., 2007; Wallace et al., 2004). This bias depends on target separation (figure 2C), as well as the posterior probability of a single cause (), and the causal inference strategy used for both the unity judgement and localization pieces of the task. If subjects are not performing causal inference at all (e.g. if they are simply probabilistically choosing to make one or two saccades, or never fusing the two stimuli into a single percept), then there will be no influence of target separation on the two-saccade localization biases (figure 2C, green line).

We compare the predictions of various models fit to both human and monkey subjects, first for the **unity judgement case**, then for the **localization case**, and finally for the **joint condition** (which more exactly represents the task subjects are being asked to perform) figs 3-5



**Figure 3:** *Unity judgement as a function of target disparity* Human subjects (n=7, left) and monkey subjects (n=2, center, right) demonstrate a pronounced preference for making one saccade when targets are close together, rather than well separated. These judgements are well fit by a Bayesian model of causal inference (green curves). this figure will be adjusted to show individual days pooled together for monkeys (n=7) instead of st dev across conditions as is currently shown.

In order to determine whether subjects are actually performing causal inference in our task, we begin by analyzing the explicit portion of the response: whether the subject made one or two saccades. We found that subjects were much more likely to make one saccade than two when the targets were coincident or close together than when they were well separated [need to do stats on this, ANOVA]. This means that the observers were not performing pure fusion (always integrating stimuli), nor pure segregation (always treating the stimuli as independent), but instead adopting some sort of causal inference strategy that is dependent on target separation. These responses are well described by an ideal observer model performing Bayesian causal inference (figure 3, green line). Importantly, humans and monkeys show qualitatively similar performance on this piece of the task (left panel vs. central and right panels). This indicates that monkeys are able to understand and perform this component of the task and have behavior that is also well described by causal inference models thus far only applied to human behavior.

We find that both humans and monkeys are well fit by models of causal inference which have the following features:xxx

Some differences are apparent between human and monkey subjects, in particular it seems like monkeys have more pronounced bias. These might be associated with learned priors over actual stimulus distributions.

**Discussion – We demonstrated that monkeys can do this task, and have some speculation about neural implementation**

Here we demonstrate that monkeys can be trained to perform a multisensory causal inference task, and do so in a manner consistent with human models of Bayesian observers.

In particular, these results are consistent with xxx result from the human literature, and extend that result to apply to macaque monkeys as well.

Some advantages of this dual task design are that it allows for rapid data collection that simultaneously captures implicit and explicit features of causal inference. This is critical for single unit recording as it is practically impossible to combine data across multiple recording sessions and ensure that the same units are being recorded.

This work is important because it will allow us to link our understanding of multisensory processing at the neuronal level with our deepening understanding of causal inference at the behavioral level. For instance we might make these predictions for what is happening at the neural level…

**Methods**

**Behavioral paradigm**

We created a novelmultisensory task closely related to tasks commonly used in the literature [refs]. This paradigm uses a dual task design, where subjects are reporting both a causal judgement (one or two targets, explicit causal inference) and the target locations (implicit causal inference) on every trial.

Subjects were seated in an anechoic chamber at a distance of XXX m from a row of speakers and LEDs located on the horizontal plane. Eye movements were monitored via magnetic eye coil (Riverbend) or video eye tracker (Eyelink XXX). While fixating at a central point, subjects were presented with either a light (green LED), sound (white noise), or both at one of 8 visual (+- 6-24 degrees in 6 degree increments) or 4 auditory (+- 6 and 24 degrees) locations. Targets were paired such that each combination of ipsilateral pairs was used (8 pairs per side, for 16 pairs), plus 4 contralateral pairs (+- 12 degrees visual paired with -+ 6 or 24 degrees auditory) for a total of 20 dual conditions. After a brief delay (600-900 ms) the fixation light was extinguished and subjects reported percepts by making saccades to the perceived stimulus location and then maintaining fixation at that target location. On conditions with multiple targets, subjects were required to make sequential saccades to each target in any order. The timing of the task was such that subjects must make both saccades in rapid succession, and so cannot adopt a strategy of waiting until the reward is delivered (or not) before making a decision about the second saccade.

**Trial filtration and saccade detection**

Trials are included as long as the subject held fixation through the go cue, and then made at least one saccade, **without enforcing any restrictions on saccade accuracy**. For multi-stimulus trials, trials which ended less than 600 ms after the go cue (the minimum duration for a successfully completed trial, see timing section [xxx]) were also excluded. This was done to minimize the number of trials which ended before the subject’s full response could be reported, and ensure that single saccades were indicative of a unified percept rather than a lapse.

Saccades were defined as any eye movement exceeding 50 degrees per second and followed by at least 30 ms of very little eye movement (max velocity <25 deg/s). Saccades of less than 3 degrees were considered corrective [XXX] and were not included as responses in subsequent analyses.

**Behavioral modeling**

We implemented a class of causal inference models which is common in human behavioral multisensory research. These models arbitrate between two sensory processing strategies. The first strategy treats sensory stimuli as completely independent, amounting to unisensory estimation of the parameter of interest (in this case, location of the source) for each stimulus. The second implements the established maximum-likelihood form of cue integration, which has been shown to provide excellent descriptions of human behavior in conditions where the disparity between multisensory cues is small or the cues are mandatorily fused [refs]. Different models of causal inference will then combine these two estimates according to specific rules, resulting in predictions that can be compared with behavior in our task.

Below we will briefly describe the important components of our models, and refer interested readers to [ref] for a much more thorough treatment of this class of models. We begin by describing the cases for location estimation under given causal assumptions (one or two cases), and then describe how these estimates are combined according to different causal inference strategies to produce both judgements about number of targets (unity task) and location of stimulus source(s).

*Location inference*

For all stimuli, internal representations are assumed to be corrupted by Gaussian noise, such that with the term denoting the actual location of the source of the respective stimulus and the term reflecting the sensory standard deviation (a free parameter). Estimates about stimulus locations for a given causal structure (c=1, common cause, eq 1; c=2, independent causes, eq 2) and internal representation can be computed via Bayes rule:

(1)

(2)

Where for the c=1 case the source is assumed to be the same for both the auditory and visual stimuli.

*Location prior*

The subject is assumed to have some prior over possible stimulus locations. A common choice in this type of model is to assume that the subjects have an independent, identical prior over both sensory stimuli,

(3)

which for the two cause case becomes,

(4)

where is the mean of the prior (often taken as 0) and is the prior standard deviation. This prior induces a compressive bias which is compatible with many psychophysical results. This prior has an additional advantage in that, because the prior and sensory likelihoods are Gaussian, the posterior is also Gaussian and can be computed analytically.

An alternative is that the subject may learn the actual stimulus distribution, rather than relying on this type of naïve compressive prior. This is particularly relevant for the monkey subjects who, unlike humans, will experience tens of thousands of trials over the course of training and experimentation. For this task we modeled the prior as either a uniform distribution over the exact target locations (perfectly empirical prior) or as a mixture of normals which grossly captures the task design of targets being concentrated either to the left or the right of fixation (heuristic empirical prior).

*Causal inference strategies*

The choice of causal inference strategy determines how the observer model decides between the c=1 and c=2 cases when presented with sensory stimuli. In general, this choice can follow either Bayesian principles, non-Bayesian heuristics, or strategies which do not actually implement causal inference at all (for example, forced fusion). There is considerable behavioral work exploring the relative merits of both Bayesian and heuristic forms of causal inference in humans, which is outside the scope of this paper. Instead we present **two(?)** forms of Bayesian causal inference common to the field, and contrast these with a null model which does not perform causal inference at all. This might need work, not really sure what models I’m talking about comparing it to.

A Bayesian strategy will compute the posterior probability of the c=1 and c=2 cause cases, given sensory information, as follows,

(5)

Where p(c) reflects the prior probability of a common cause , which is left as a free parameter. Because there are only two possibilities for causal state in this paradigm, this can be written as

(6)

The sensory likelihoods will depend on the choice of prior in the previous section, according to

(7)

(8)

For the simple normal prior these can be solved analytically, but for other forms of prior numerical integration is required. To ensure fairness during the model comparison steps, all likelihoods are computed using the same method (numerical integration).

*Decision rule*

The final component of the observer model is the decision rule used to choose responses based on the above probability distributions. There are three common choices for this decision rule: model selection, probability matching, and model averaging (for the unity judgement task, only model selection and probability matching are relevant). Below we give an overview of these decision strategies and how they relate to subject responses in the unity judgement and localization components of the task.

For the unity judgement task, model selection corresponds with reporting whichever causal scenario has the highest posterior probability,

(9)

where represents the lapse rate (where the subject randomly makes a response), and is the Iverson bracket which is 1 when the statement inside is true and 0 otherwise.

For the probability matching case, the observer randomly chooses unity at a rate consistent with the posterior probability of the causal inference. In this case the probability of reporting unity is exactly equal to the posterior probability , except for the additional lapse parameter described above.

For the localization component of the task, subjects must arbitrate between the two potential models of location conditioned on number of causes (eqs. 3 and 4). This amounts to reweighting the two estimates according to some weight function that is dependent on the sensory response,

(10)

where defines the decision weight applied to the c=1 condition.

For the model selection strategy, this weight is equivalent to eq. 9, such that the weight applied to the c=1 condition is 1 when that is the most likely causal scenario and 0 otherwise. A similar rule applies for the probability matching scenario, except that the weights are randomly set to 1 with probability and 0 otherwise. For model averaging, the weight is equivalent to the posterior probability , such that the expression in eq. 10 becomes equivalent to the optimal observer model commonly used in the literature [ref][may need to revisit this section for clarity].

*Comparing with behavioral data*

The above response estimates are dependent on internal variables and which are not accessible to the experimenter. To get distributions that can be compared with data, eq. 10 must be marginalized across the internal variables,

(11)

We compute this distribution using numerical integration for each of the 20 combinations of visual and auditory targets.

**Model fitting**

Models are fit using a maximum likelihood approach to determine the set of parameters which best explains the provided data. This is accomplished using [approach][ref]. Models are validated using a k-fold cross validation strategy, holding out 1/5th of trials for testing and training on the other 4/5th of the data. Total model likelihood used for model comparison is computed by summing likelihoods across all 5 folds of test datasets.