Behavioral causal inference model technical report

For outlining the specific modeling steps that have been taken for translating the causal inference model of multisensory perception (initially proposed by kording et al 2007 and others) into a primate saccade paradigm. This is not meant to approximate a useful draft but is meant instead to serve as a jumping off point.

Key points I want to focus on in this paper:

1. does CI provide a good explanation of behavior when decision making is explicit and natural
2. Is CI then involved in perception of target locations in an optimal way, or is there some other non- optimal solution like model selection between integrate and segregate happening?

**Short term needs:**

* ~~Finish methods section up to current state~~
  + ~~Paradigm~~
  + ~~Generative model~~
  + ~~Likelihood calculation~~
  + ~~Model comparison~~
* ~~Results section~~
  + ~~Does the model do a good job capturing behavior qualitatively~~
  + ~~What are some issues/discrepancies (collect these in a list on google doc as well)~~
  + ~~Which model is the best~~
* ~~Add some figures, especially displaying the model fits~~
  + ~~Model fits of saccade distributions~~
  + ~~Model fits of n saccades~~
  + ~~Model comparison results~~
* ~~Manuscript plan/story~~
* ~~List of experiments to run~~
* ~~List of analysis improvements~~
* ~~Update google doc~~
* ~~Prioritize~~

**Long term needs:**

* Intro and conclusion
* Replace equations with non-pictures
* Rework into a story
* Formatting
* Compare CI and alternative models

**Introduction**

-contrast with integration condition

- set up CI models as a useful step forward in understanding multisensory perception and perceptual decision making

- motivate use of monkeys for addressing this question

- benefits of using saccades as a report

**Methods**

**Behavioral Paradigm**

Subjects are 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. While fixating at a central point, subjects were presented with either a light (green LED), sound (white noise), or both at one of 10 visual (+- 6-30 degrees in 6 degree increments) or 4 auditory (+- 6 and 24 degrees) locations. After a brief delay (600-900 ms) Subjects reported percepts by making saccades to the perceived location in space and then maintaining fixation at that target location. On conditions with multiple targets, subjects are required to make sequential saccades to each target in any order. The timing of the task is such that 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. This means that subjects are making both an explicit causal inference judgement (number of saccades) and an implicit judgement (location of fused percept) on each multisensory trial.

**Saccade detection**

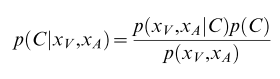
Saccades were defined as any eye movement exceeding 50 degrees per second and followed by at least 20 ms of very little eye movement (max velocity <25 deg/s). Endpoints were calculated by taking the average over the 15 samples immediately following the termination of the saccade. This 20 ms pause is much shorter than what the monkeys are required to do in order to get a reward in the paradigm (100 ms pause minimum at target location) but is enough to ensure that they actually stopped rather than just slowed their eyes down. I feel that this is more reflective of the actual behavioral report, as enforcing the 100ms pause after the fact undercounts saccades, especially the cheat left ones. For this analysis saccades are only included if the endpoint occurred between the go cue and the end of a trial (either reward or failure) plus a 100ms buffer (to account for saccades which were initiated before the end of the trial).

**Perceptual Modeling (I definitely need to review the details of this)**

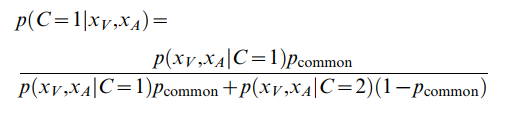
We implemented a form of Bayesian causal inference as first described by Kording et at (XXX). This approach uses a generative model which is defined as follows. First, the model assumes that the one (C=1) or two (C=2) cause conditions are determined by drawing from a binomial distribution with the common source prior P(C=1) = pcommon. The model also assumes that there exists a normal prior distribution of possible target locations described by N(mup, sigp) this might be changed in the future. In this iteration the value mup = 0, reflecting an distribution centered on the midline. When in the C=1 condition, a single target location is drawn from the prior distribution (SAV) and used to generate estimates of both auditory (xA) and visual (xV) percepts by sampling from a normal distribution centered on SAV with variances specified by sigA^2 (currently there are two sigA parameters, one for close targets and one for distant targets, this might change) and sigV^2 respectively. These draws are meant to reflect sensory noise corrupting the target estimate. In the C=2 condition, two target locations are drawn independently (SA and SV) and estimates xA and xV are then drawn from these different locations. This leaves the model with four free parameters: the prior on common cause (pcommon), the centrality bias (sigp), and the sensory noise (sigA and sigV).

**Estimating probability on common cause:**

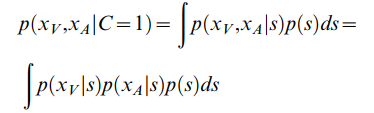
The causal inference model implements a form of Bayesian causal inference that estimates the likelihood that two stimuli originate in the same spatial location, given a separate estimate of each modality. This has the form:



Where C =1 reflects a common cause and C=2 reflects separate causes. Because there can only be one or two causes the probability for these must sum to one. We can simplify this to

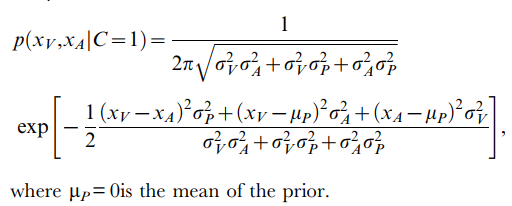


Where

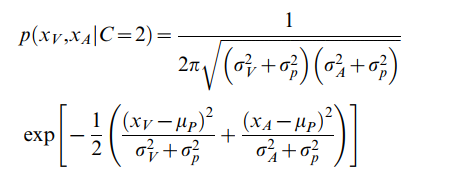


Here *s* is the single source that generates both xa and xv. In the two cause case there is an sa and sv term that are independent from one another, but drawn from the same underlying stimulus distribution.

Because in the above integral all three factors are gaussians, there is an analytic solution to the integral. Namely:



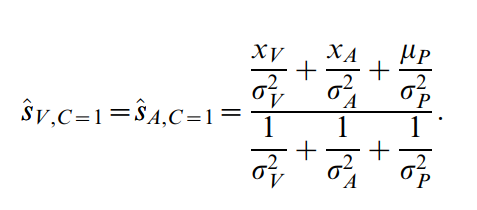
Similarly we can do this for the two cause case, eventually ending up with,

.

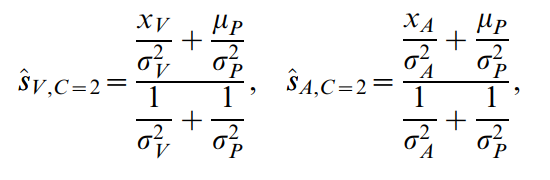
This allows for comparison with the experimental data, specifically for the number of saccades the monkey will make for a given stimulus pairing on average. This assumes the monkey always makes one saccade when p(C=1|xv,xa)>0.5, but this threshold might be shifted in the future to account for behavioral biases (for instance, might expect the monkey to be biased towards one saccade because it is less challenging).

**Optimally estimating location**

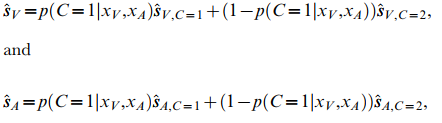
There are two strategies for optimally estimating the location of *s*, the source of the stimuli, conditional on whether C=1 or C=2. For C=1, it is a problem of optimal Bayesian integration, which takes the form:



While for multiple sources there should be two estimates of *s,* one for each modality, and with no interactions between modalities. This takes the form:



These are the estimates assuming external knowledge about the existence of one or two targets. Therefore in order to compare with real data (which will have a mix of one and two saccade responses, indicating a subjective perception of one or two targets), it is necessary to combine the estimates according to some judgement of causal inference. This was done using the weighted averaging strategy that has been reported as the most common perceptual strategy used by humans in a similar task (rohe ref). This took the form:



Effectively this strategy reweights each of the optimal sensory estimates according to their posterior on common cause…

An alternative approach would be to use a model selection strategy, where the subject makes a causal inference judgement and then specifically chooses either the one cause or two cause estimation strategy. This is an area of future investigation.

**Model fitting**

The model is fit by maximizing the likelihood (minimizing the log likelihood) of the actual saccade distributions under the model. This was done using the fminsearch function in matlab, with five free parameters. These were the prior on common cause (p(C)), visual sensory variance (σV), auditory variance (σAn for near targets and σAf for distant targets), and centrality bias (σc). Each of these parameters was given an initial starting point determined by a prior grid search. Likelihood for each set of parameters was obtained by calculating the actual likelihood of a set of 1000 saccades taken from real behavioral data for the given target condition, under the calculated probability distribution (I don’t know if this is right).

Likelihood was combined across all target pair conditions for a given parameter set, which enforced the assumption that the same parameter values were used for each condition (rather than each target pair having its own variance parameter, for example). In order to compensate for overweighting of dual saccade trials in the model fitting step (because these trials would contribute two saccades to the dataset, rather than the single contribution of unified trials), all trials were subsampled such that they only contributed a single saccade to the final distribution. An alternative approach might be to explicitly model the number of saccades as a feature that must be fit, which would penalize the generation of extra saccades when two causes are improperly inferred.

This procedure was repeated for simplified models that assume the targets are always at the same location (integrate model) and models that assume there are always two locations (segregated model). These models have identical structure to the causal inference model where p(C) is set to 1 and 0, respectively.

After models were fit, likelihood was calculated using a “held out” dataset of XXX trials. Models were compared against one another using AIC, which corrects for the larger number of free parameters in the CI model.

… somewhere above should insert that the “true” locations (Sv, Sa and Sav) are fed in as inputs to the model. I think xa and xv are taken as the mean samples from the unisensory responses, but I’m not sure about this. This is probably a bug in the code and should be fixed or clarified.

**Model Comparison**

Models were compared by calculating the likelihood on a held out set of 1000 trials per condition taken randomly from a larger pool collapsed across behavioral days. Goodness of fit was compared using AIC, which punishes extra model parameters.

**Deviations from previous modeling (do this second)**

List of changes:

-Saccades (analog signal)

- different likelihood calculation (because using analog)

- extra aud variance term

- holding out data

The biggest difference between this paper and any previous work on perceptual causal inference is the reporting of both auditory and visual perceived locations on all trials. This is beneficial because it incorporates both an explicit causal inference report (number of saccades) as well as an implicit one (influence of non-localized modality on saccade endpoint). However, this requires several adjustments in order to disentangle sensory and motor effects.

One of the major differences from previous models (kording) is the inclusion of a second auditory variance term. This was done because the variance on very lateral auditory targets was much lower than the variance on very medial targets. There are two reasons for this difference, the first is that the auditory discrimination is more challenging for the middle targets (12 degrees from nearest aud target, 6 deg off fixation) than on the lateral targets (18 degrees from nearest aud target, 24 degrees from fixation). The second is that the monkey has been trained to make a saccade of at least 6 degrees to either the left or right on every trial. This biases the monkey towards making large saccades, even when the target is actually “perceived” near the midline. We think it’s possible that when a monkey is making a difficult discrimination (i.e. is this target to the left or right of center?) they bias their reports to locations either on the left or right of the fixation light, rather than the true perceptual location near the midline. This would result in an overdispersion of responses, with a bimodal distribution rather than a purely unimodal one (XXX is this the case in the data? Also could be tested by changing the prior from a central one to a bimodal one).

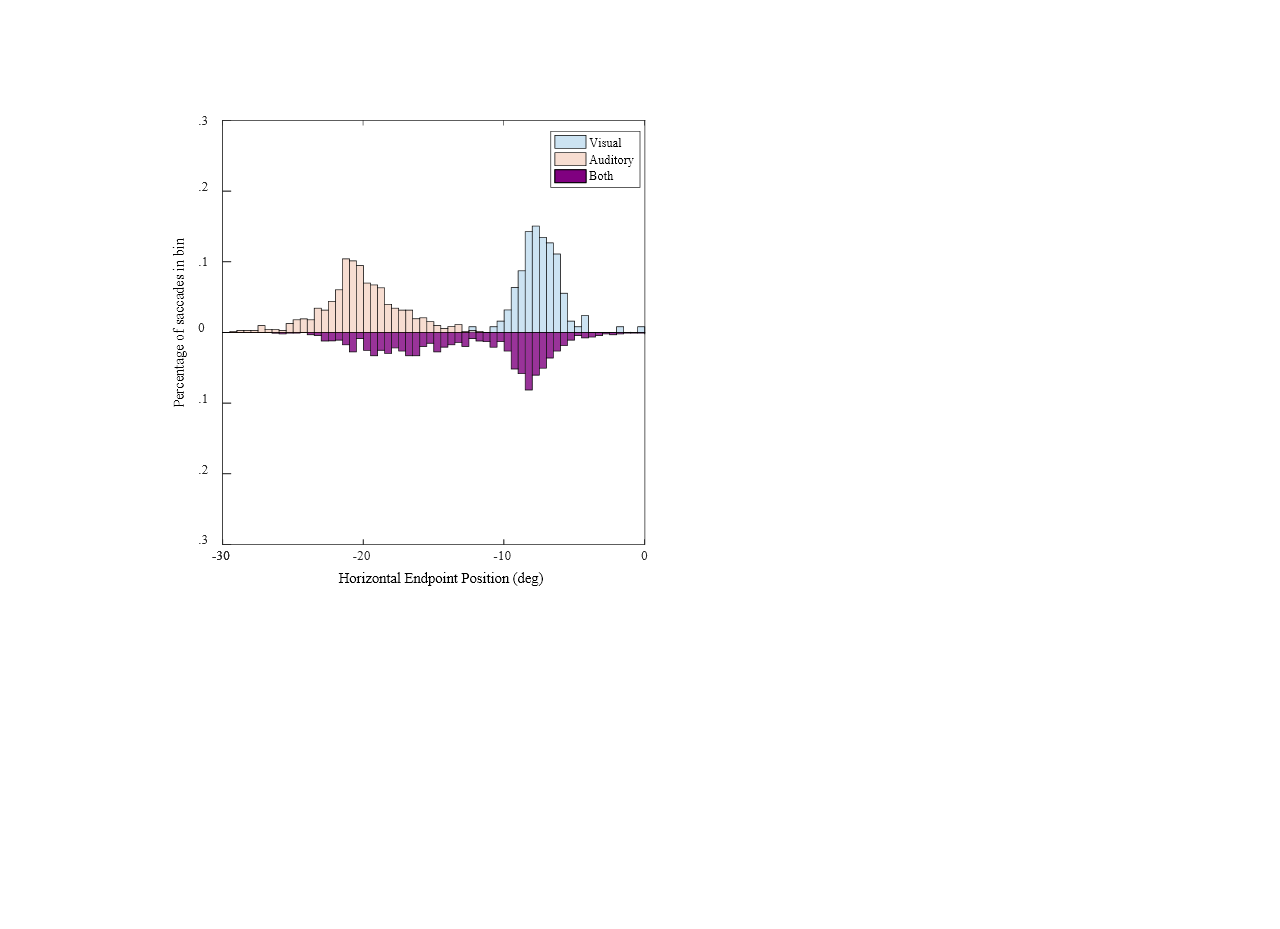
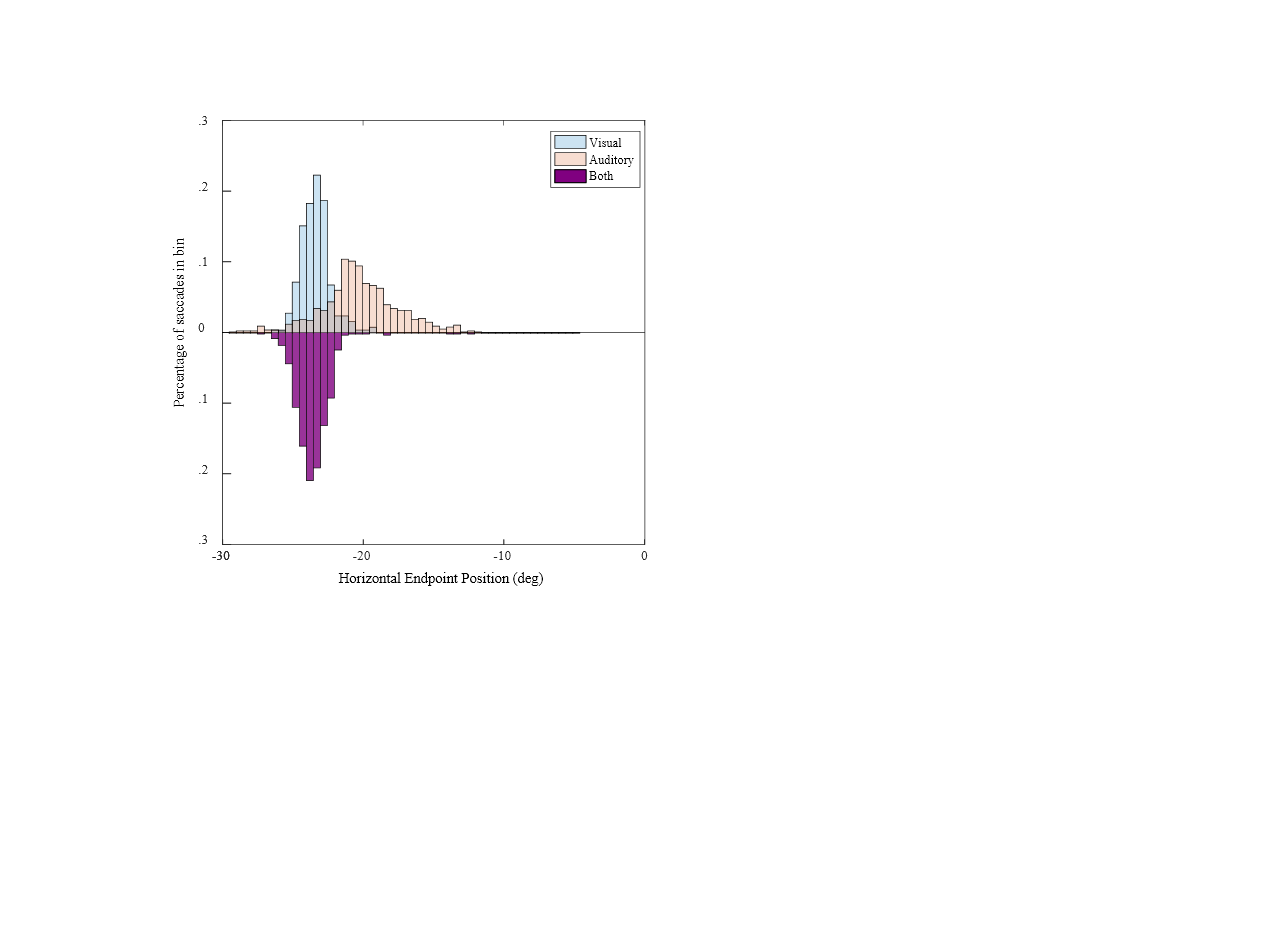
**Results**

* Does the model capture behavior well
* How does the model compare with other possibilities
* What are some things we can pull out from looking at the model fits (accuracy, bias, priors)
* How do monkeys and humans compare in this paradigm
* What does this tell us about CI as a strategy used in the brain

**Monkeys are able to localize both a light and sound in multimodal trials**

Monkeys could be trained to sequentially report both an auditory and visual stimulus, even when presented at the same time. As expected under visual capture, monkeys were much more likely to make a single, visual-cue-dominated saccade in conditions where the auditory and visual stimuli were close together (within 12 degrees, figure 1 left) and much more likely to make independent saccades to both target locations when the targets were well separated (figure 1, right). This relationship between target separation and number of saccades is shown in figure 2C (might want to make this a separate figure from the one showing the modeling results).

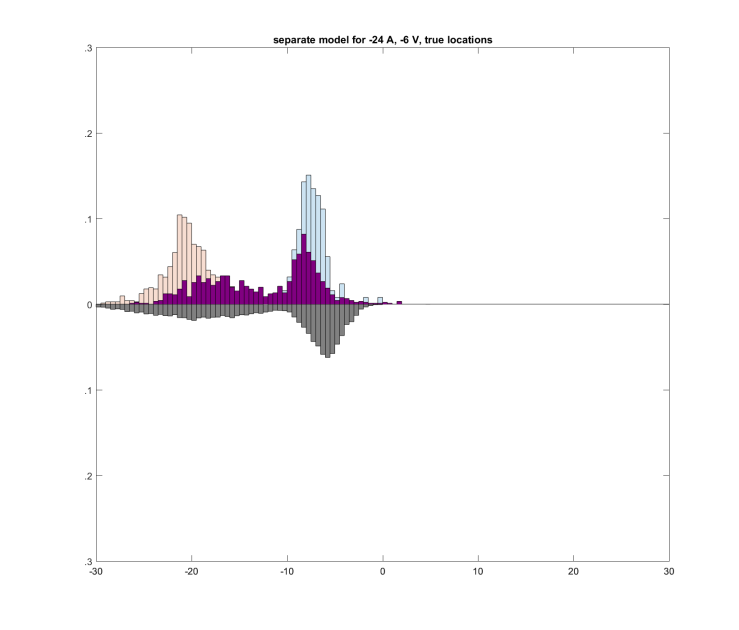
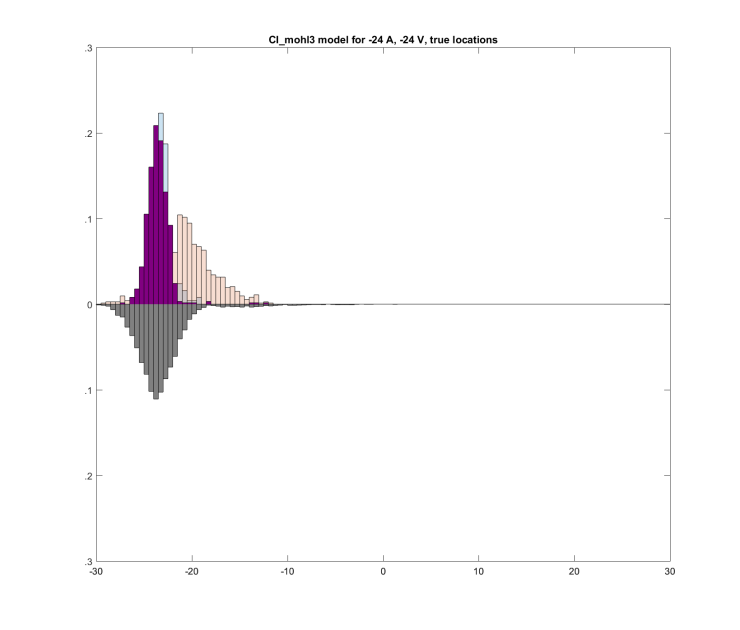
Monkeys did not show a strong preference for order between stimuli. Although first target selection was not random within conditions (test for this) monkeys appeared to switch preferences between conditions, possibly preferring saccade patterns that were easier to generate from a motor perspective rather than preferring one stimulus over another.

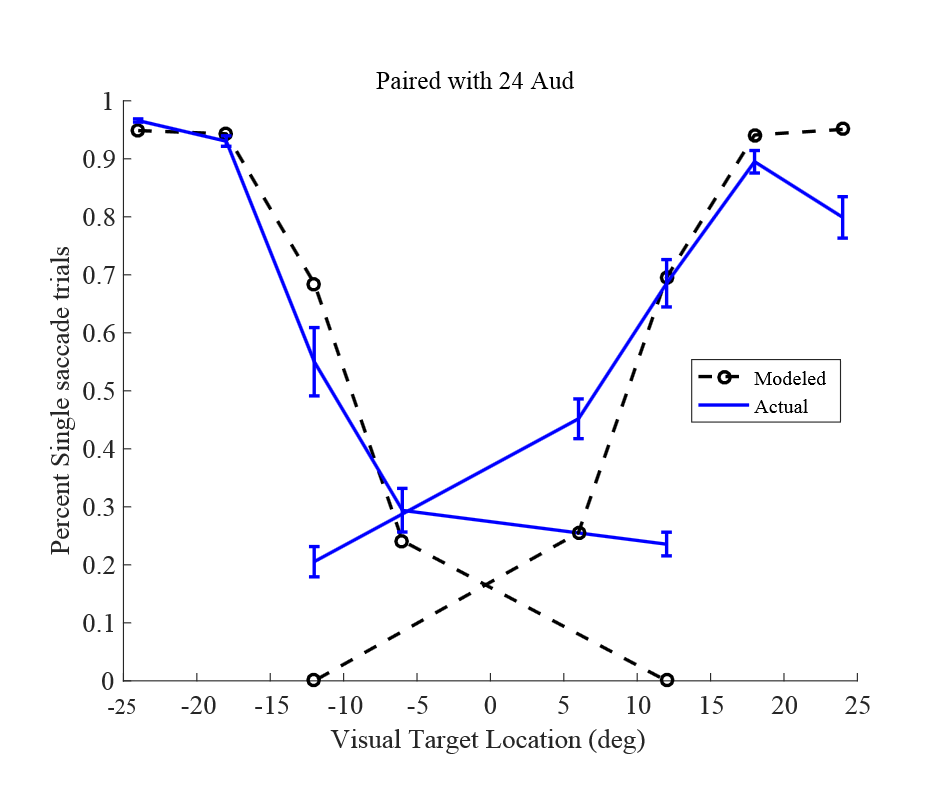


**Figure 1:** *Integrated (left) vs segregated (right) target configuration. Monkeys are able to report of a fused percept (left, purple) when the unisensory percepts (visual, blue; and auditory, red) have significant overlap. When the two distributions are easily separable, monkeys make two saccades approximately matching those made in the single modality condition (right, purple).*

**CI model qualitatively captures monkey behavior**

In line with previous studies, the causal inference model was able to grossly capture response locations of monkeys in both single saccade (figure 2A) and multi saccade (figure 2B) conditions, including the presence of both average and winner take all saccades in the multi-saccade case (more explicit about this). Additionally the model was able to capture the relationship between target separation and number of saccades, which was never explicitly fit in the model (figure 2C).



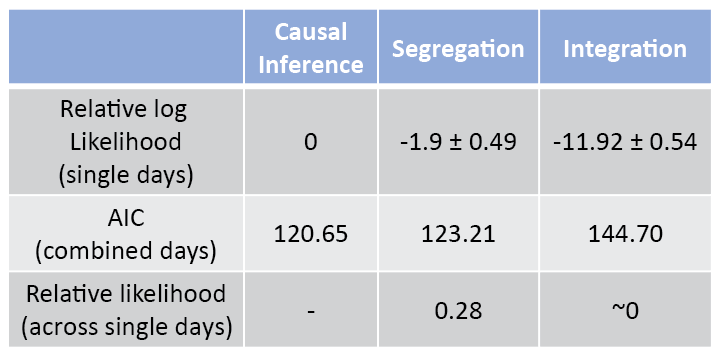


**Figure 2:** *results of CI model fitting* (A) Model fits for an overlapping visual and auditory stimulus (black). (B) model fits for a well separable stimulus pair. Note the bimodality of the modeled distribution. (C) number of saccades produced by generative model compared with actual behavior. This feature was not fit by the model but instead falls out automatically from the causal inference judgement

Qualitatively, these results suggest that monkeys are both integrating and segregating stimuli in a fashion that is consistent with causal inference. Monkeys display a mix of integration, visual capture of auditory stimuli, and complete segregation, and these behaviors depend on the degree of separation between targets. However, some important deviations are obvious. First the model greatly overestimates the variance of the combined stimulus condition in both the integrate and segregate conditions. This can be seen in the over-dispersion of grey vs purple distributions in figure 2. The most likely explanation for this is the presence of erratic saccade behavior on a subset of trials (see examples in -6a-12v, where purple distribution has a very long tail). The way that models are fit disproportionately emphasizes these points, and getting rid of them is an obvious way to improve the current model. Another major difference is the disagreement between optimal and observed behavior for explicit CI as reported by number of saccades. An optimal observer would have conditions in which 2 saccades were made 100% of the time, whereas the monkey actually makes only one saccade in at least 20% of trials regardless of stimulus configuration (see figure 2 bottom, dashed vs solid blue line). This is likely because of a behavioral lapse, in which the monkey did not initiate the second saccade quickly enough or was not paying close attention on that trial. This could potentially be compensated for by including a lapse parameter in the model fits.

**Quantitatively, CI model performs better than either pure integration or pure segregation models**

In order to determine if the CI model performed better than either of the potential simpler models considered (pure integration of stimuli or pure unisensory target estimation), a direct comparison using AIC was performed (Table 1). From this it can be seen that the CI model provides only a modest improvement over the complete segregation model (which assumes that the monkey always makes two saccades, one to each target location). This is in line with the qualitative observation that the distributions generated by the CI and segregation models appear very similar, with only a modest ‘filling in’ between two closely spaced unisensory distributions in the case of the CI model. However this model comparison only the likelihood of the observed saccades under each model, and does not take into consideration the explicit report of one vs two saccades. When comparing the number of saccades by target separation, it is clear that the CI model provides a better explanation for what is actually going on (figure 2 bottom). But I don’t really have a way to quantify this yet.



**Table 1:** *Results of model comparison against other potential strategies* Models were compared using AIC, which punishes models with extra parameters. An AIC difference of 2 is equivalent to adding an additional parameter.

**Shortcomings of model**

**…** some of the things I would consider shortcomings are actually in previous sections of the results…

It is worth noting that because the visual and auditory stimuli are both well above perceptual threshold, the effects of multisensory integration are expected to be relatively weak (due to the principle of inverse effectiveness ref). Therefore the explicit causal inference, as reported through the number of saccades, is more informative than the degree of implicit causal inference (i.e., the amount of visual capture in the single saccade trials) when judging the differences between bound and separate percepts.

**Conclusion**