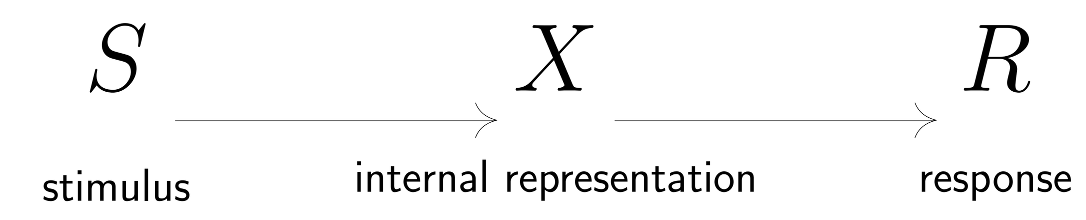
Match identity explains the response process in continuous orientation recall: a Bayesian inference framework

Kira Düsterwald[1], Peter Vincent[2], Athena Akrami[2] and Maneesh Sahani[1]

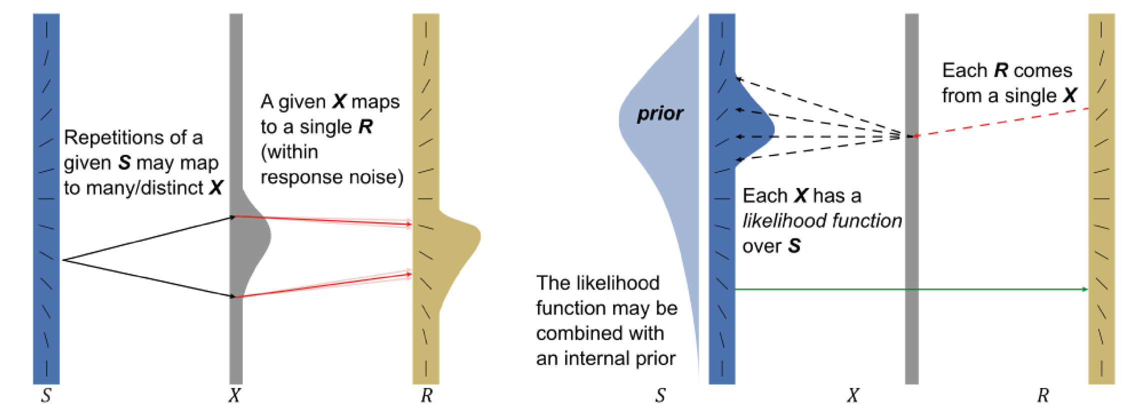
[1] Gatsby Computational Neuroscience Unit and [2] Sainsbury Wellcome Centre, University College London

Introduction

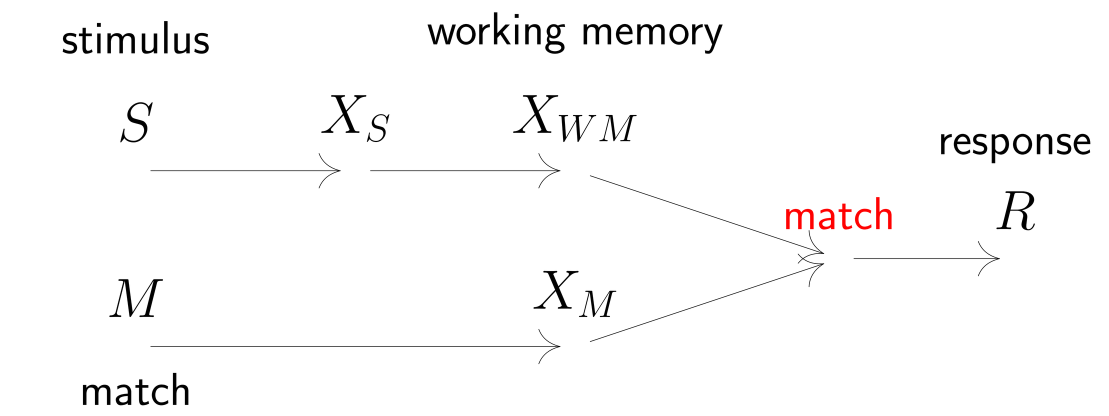
An organism's reproduction R of a stimulus S is not exactly S; there is systematic behavioural variability.



1. How do humans use the posterior to make a response?
   - Deterministic *e.g. Mean, mode or median*
   - Posterior sampling



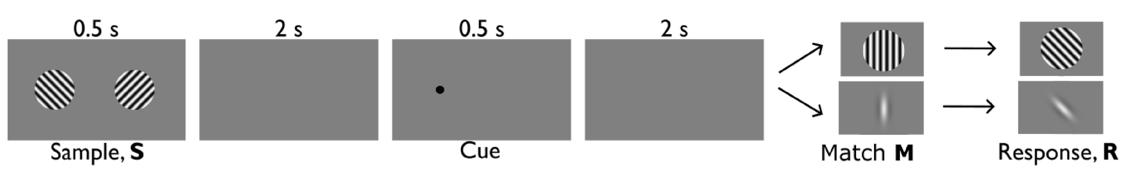
2. Consider including a participant's reporting tool, the **match**.



If distributional beliefs are all you need, changing stimulus features in the match should not matter.

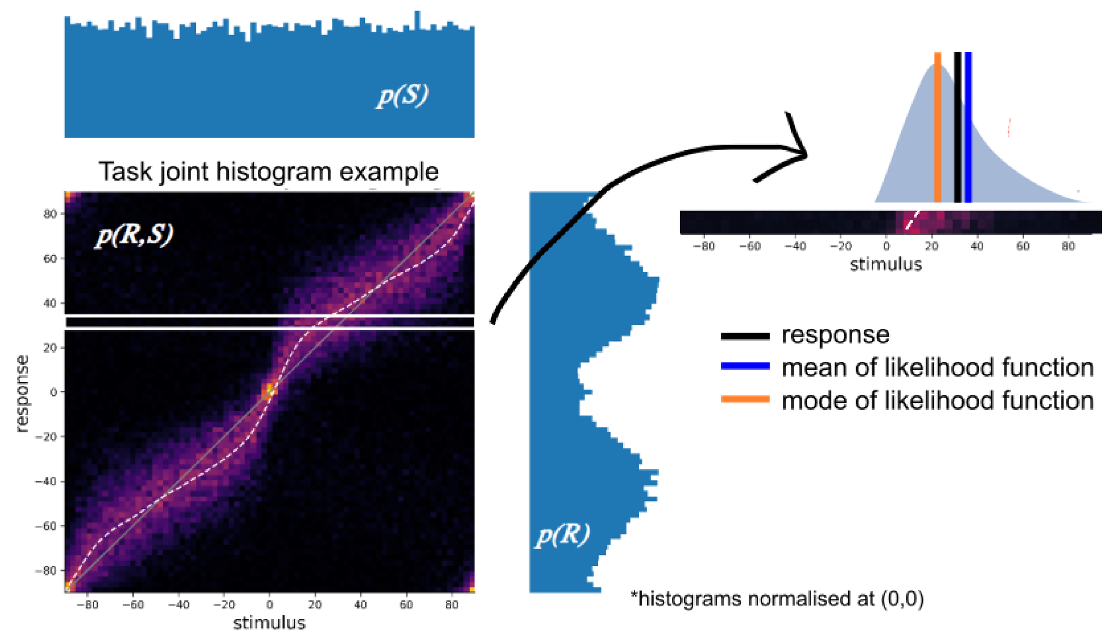
Visual orientation reconstruction task with different matches

Participants are cued to reconstruct one of two remembered visual orientation gratings ($S \sim \mathcal{U}(-90, 90]$) by rotating a **match**, which can be a grating or Gabor.



This task and similar ones produce the well-known picture of bias away from cardinal (horizontal and vertical) orientations. Various models aim to explain this bias, e.g via efficient coding (e.g. Taylor and Bays, 2018), but cannot account fully for how participants choose responses.

We use ideas from Bayesian inference to create models for psychophysical (behavioural) data.

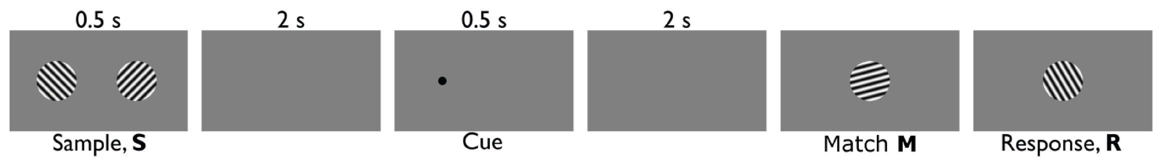


We directly recover **likelihood functions** from the joint sample-response histogram, which participants may combine with an **internal prior** to create a **posterior** – their response.

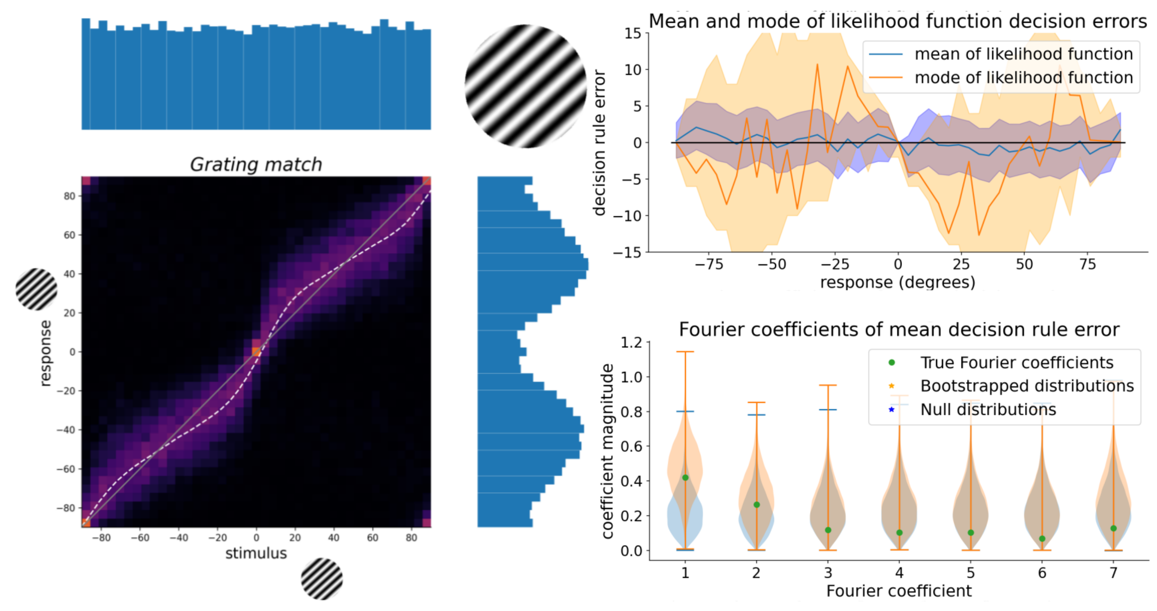
What statistic do participants use from the likelihood function (histogram rows) to decide their response? We can compute the mean as the circular mean and the mode using a kernel density estimate with von Mises kernels.

Expectations of distributional beliefs can explain participant responses

Responses with a grating match stimulus correspond to the **mean** of the likelihood function rather than the mode.



We fail to find significant Fourier coefficient magnitudes for the error between likelihood function mean and true response, comparing the bootstrapped distribution to a constructed null 0-error distribution – the error is consistent with a constant, 0.

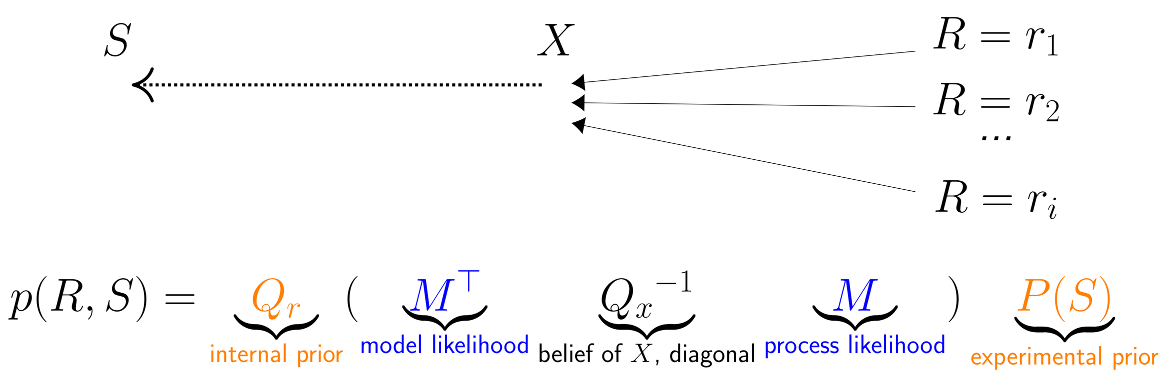


We replicate agreement with the mean decision rule when the stimulus is a Gabor and the match a grating: it is match-dependent.

Responses seem to be distributionally accurate in expectation, employing a uniform (flat) internal learned prior, in contradiction with statistical learning theory.

Posterior sampling cannot explain the data

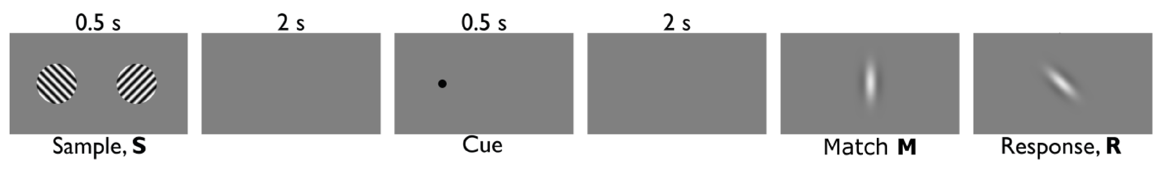
Additionally, posterior sampling cannot explain the data.

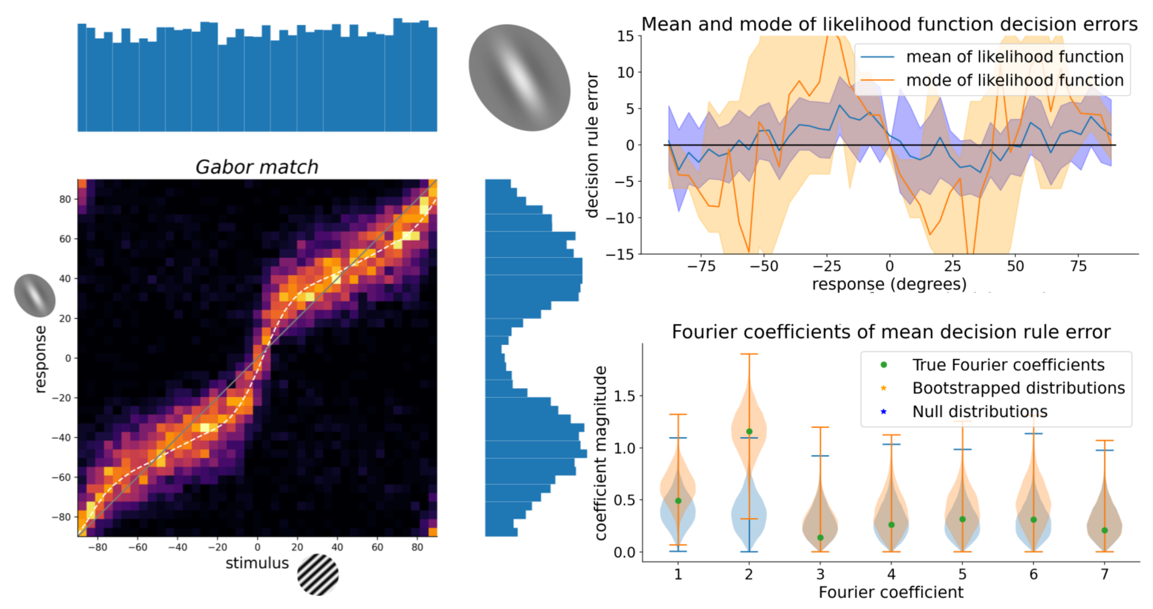


Random sampling is symmetric up to how S are produced, so the joint histogram should be symmetric for a uniform prior, yet it isn't. No prior Q_r can be fitted that reconciles the result.

Surprisingly, results depend on match identity

The mean of the likelihood function does not explain responses when the match is a Gabor.





We replicate violation of the mean decision rule when the stimulus and match are both Gabors: it is match-dependent.

Responses are pushed away from cardinal relative to the likelihood function mode or mean.

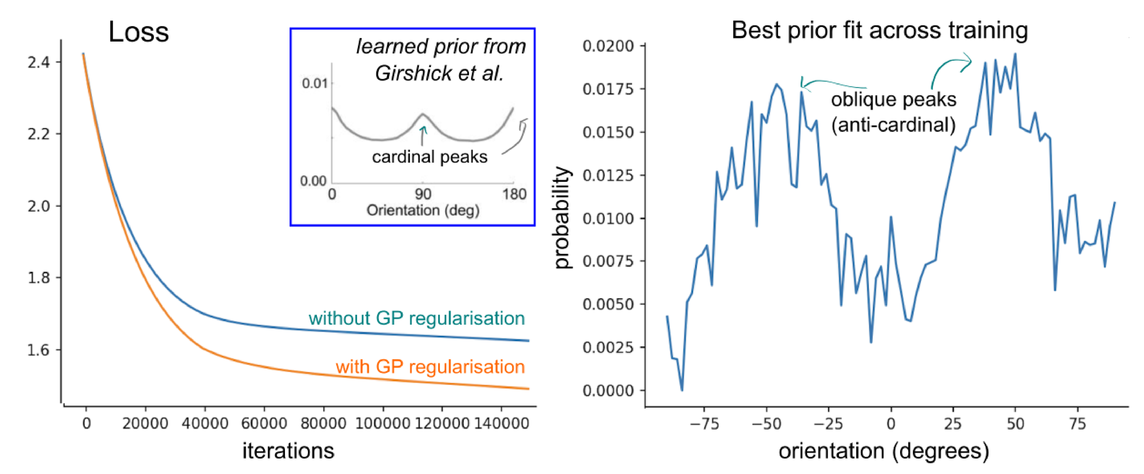
The optimally fitted prior is anti-cardinal, contradicting literature

What if we employ a non-uniform prior on the Gabor match likelihood functions, then take the mean of the posterior?

Given $r = \mathbb{E}\left[\int e^{iS} F(S|r) p_{internal}(S) dS\right] \forall r$, where $F(S|r)$ is the likelihood function, we fit an optimal learned prior $p_{internal}$ using gradient descent. In particular, given the loss we find $p_{internal} = \arg\min_{p_{int}} L(p_{int}(S))$, where

$$L(p_{int}(S)) = \sum_r \left(\cos(r) - \cos(F(S|r)p_{int}(S))\right)^2 + \left(\sin(r) - \sin(F(S|r)p_{int}(S))\right)^2 + \lambda GP_{loss}$$

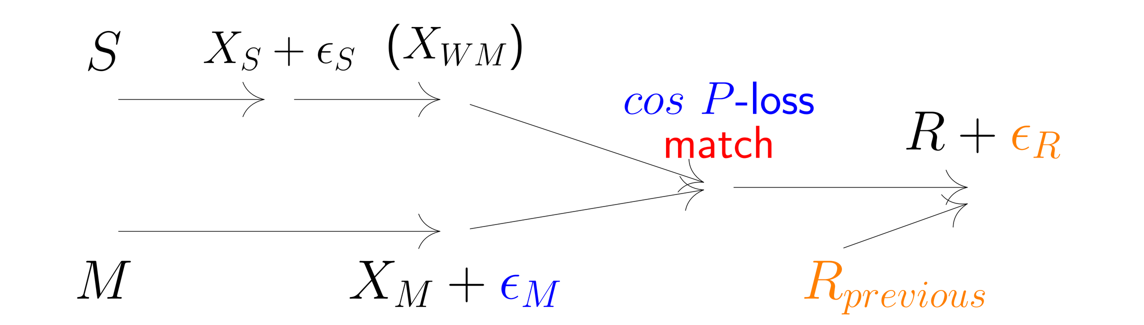
The Gaussian Process loss regularisation uses a RBF kernel – purpose is to smooth the fitted prior. We have to normalise the prior density to 1, but only at the end of the descent algorithm.



The optimal prior is anti-cardinal, contradicting previous findings based on statistical learning that priors are pro-cardinal (Girshick et al., 2010).

Which other models could account for anti-cardinal response shifts?

1. The following models cannot work:
   (a) Serial one-back effects (conditioning on most recent R orientation) biases cardinally.
   (b) Adding symmetric response noise (ϵ_R) biases cardinally.
   (c) Selectively switching strategy for stimuli and matches of the same identity, i.e. pixel-matching (De Gardelle, 2010).
2. Current work involves exploring:
   (a) **Higher-order (cosine) loss functions** between response and stimulus internal distributions (Hahn and Wei, 2022), e.g. $\cos_P(\theta_S, \theta_M) = (1 - \cos(\theta_S - \theta_M))^{\frac{P}{2}}$.
   (b) **Incorporating uncertainty** in the internal representation of the match (Mao and Stocker, 2022): optimise/grid search von Mises κs in $p(R^*|S) = \iiint p(R^*|R)p(R|X_M)p(X_M|X_{WM})p(x_{WM}|S)dRdX_MdX_{WM}$.



Conclusions

- A probabilistic ML and Bayesian inference lens and using machine learning techniques like optimisation and Gaussian processes is informative for neuroscience studies.
- A circular, continuous report dataset allows for identification of complete likelihoods.
- Surprisingly, changes in match identity affect response strategy. Match characteristics that may be important could include spatial frequency and Gaussian envelopes.
- To our knowledge, in existing models that might explain the influence of a match, responses are shifted towards cardinal (if at all). This cannot account for the anti-cardinal effect when participants rotate a Gabor match.

References

1. Taylor and Bays, *J Neurosci* 2018
2. Girshick et al., *Nat Neurosci* 2010
3. Noel et al., *PLOSBio* 2021
4. De Gardelle et al., *JoV* 2010
5. Hahn and Wei, *bioRxiv* 2022
6. Mao and Stocker, *bioRxiv* 2022

✉: kira.dusterwald.21@ucl.ac.uk