Responses in an orientation recall task are generated by taking expectations of distributional beliefs

Peter Vincent (peter.vincent.14@ucl.ac.uk)
Sainsbury Wellcome Centre, 25 Howland Street, London, UK

Athena Akrami[1] (athena.akrami@ucl.ac.uk)
Sainsbury Wellcome Centre, 25 Howland Street, London, UK

Maneesh Sahani[1] (maneesh@gatsby.ucl.ac.uk)
Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK

Abstract

The perception of sensory stimuli is frequently variable. Previous studies have shown that observers account for uncertainty arising from internal variability when they combine sensory cues, integrate sensory input with prior expectation, or select actions under externally imposed cost functions. But how does the distribution of internal uncertainty shape free perceptual report? Behavioural models have assumed that unitary percepts may reflect means, modes or samples of internal belief distributions. Here, we show that observers' reconstructions of the remembered orientation of a visual grating correspond to means of the variability-induced likelihood, not the mode or a random sample. This behaviour remains robust as either the distribution of stimuli, or the degree of internal variability, change. These observations suggest that variability arises in encoding or recall, and is accurately taken into account at the point of perception or action.

Keywords: decision making; working memory; behaviour

Task and behaviour

Human participants (n=48) were presented with a pair of circular visual gratings with orientations sampled independently from the same distribution defined on $[-\frac{\pi}{2}, \frac{\pi}{2})$. The gratings remained visible for 0.5 s, followed by a blank screen for 2 s, and then a 0.5 s-long cue (a dot at the position of one of the two gratings) to indicate which stimulus was to be recalled. After another 2 s blank, a new grating appeared in the centre of the screen. Participants rotated this grating to match the orientation of the cued stimulus, submitting a response by pressing the space bar (Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). They received no feedback. Normalised joint histograms $P(R,S)$ of cued sample orientations S and responses R are shown in Fig 1B and Fig 1E (uniform and von-Mises distribution of S, respectively). Marginal distributions of S are shown in blue and of R in yellow. The correlation between the participants' response and the uncued stimulus was close to 0. Responses displayed a characteristic pattern of bias and variance around the sample: variance was tightest near the cardinal directions, but small tilts away from the cardinal axes were systematically exaggerated in recall. Others have speculated on the origin of the bias (Taylor & Bays, 2018). Here, we looked at how the pattern of responses could reveal the decision rule adopted by the participants.

Mapping from belief to action

We use capital letters for random variables; lower case for realisations; $p(\cdot)$ for true distributions, and $q(\cdot)$ for participants' beliefs. Forms like $p(R = r \mid S)$ or $p(r \mid S)$ should be read as functions of the capitalised random variable, here ranging over the values of S for fixed r. $\mathbb{E}p$ is the (circular) mean of p.

Consider a model in which the cued stimulus S is recalled in a noisy internal representation X, which informs the response R. The row of the joint histogram in Fig 1B corresponding to

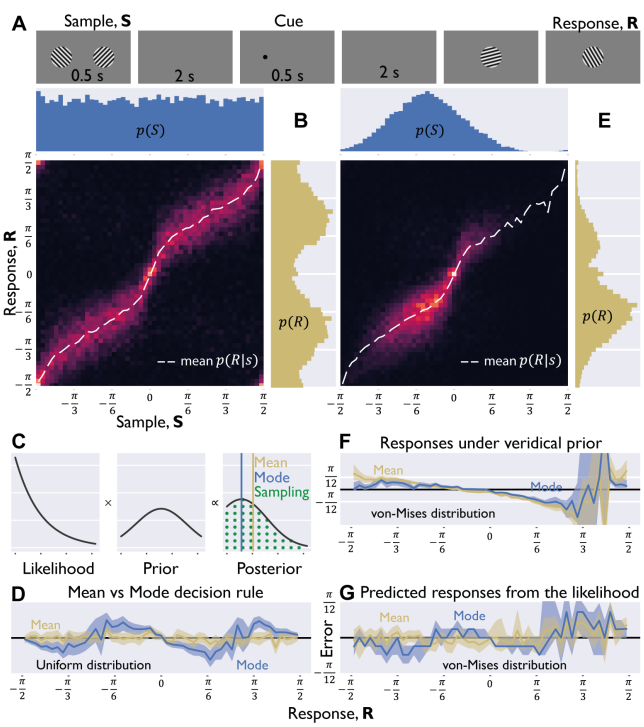

Figure 1: A - Task schematic. B,E - Joint and marginal data distributions. C - Schematic of different decision rules. D,F,G - Model errors from different decision rules.

[1]This work was jointly supervised by these authors.

response r is the likelihood $p(r \mid S)$, an unnormalised distribution over S. Although we do not have direct access to the internal variable X, we can use the pattern of these likelihood functions to test plausible relationships between X and R.

Deterministic mappings from X to R. Consider first the possibility that X maps one-to-one to R. In this case, each row $p(R = r \mid S)$ corresponds to a unique internal likelihood $p(X = x \mid S)$. We assume that participants adopt a prior $q(S)$ and choose a response based on the posterior $q(S \mid x) \propto q(S)p(x \mid S)$. The question is how this choice is made. Consider the belief $q_r(S \mid x)$ derived from the $R = r$ row of the joint histogram. A consistent choice function must be one that maps $q_r(S \mid x) \mapsto r$. We evaluated two proposed rules (Wei & Stocker, 2015) by finding the (circular) means and (smoothed) modes of the inferred distributions (also see Fig 1C). We first considered likelihoods derived from the uniform sample distribution (Fig 1B), assuming a uniform internal prior. Fig 1D shows the differences between the two statistics and the corresponding participants' response (with bootstrapped 99% confidence intervals). Responses were consistent with the means. In the second experiment (Fig 1E), we saw scant evidence that participants incorporated the non-uniform sample distribution (blue histogram). Performance throughout the experiment remained consistent, suggesting a lack of learning. Neither the means nor modes of posteriors based on the veridical prior matched behaviour (Fig 1F). By contrast, means (but not modes) of posterior distributions under an assumed uniform $q(S)$ predicted responses (Fig 1G). Note that this agreement arises despite substantial asymmetry in responses. Neither mean nor mode of $P(R \mid s)$ equals s.

Stochastic mappings from X to R. The mean consistency property also holds when mean-based percepts are corrupted by homogeneous zero-mean motor or response noise. Consider an idealised histogram $p(\hat{R}, S)$ without such noise, with $\mathbb{E}p(S|\hat{r}) = \hat{r}$. Response noise leads to rows of this idealised histogram being mixed together in the observed $p(R,S)$ with weights given by the noise distribution $p(R|\hat{R})$. As long as $\mathbb{E}p(r \mid \hat{R}) = r$, the mean-mapping property will be inherited.

Another plausible stochastic rule is posterior sampling: $R \sim q(S|X)$. Fig 2A,B shows that this model implies a core symmetry in $p(R,S)$, possibly broken if the subjective prior $q(S)$ does not match the veridical $p(S)$. Clearly the observed histogram is far from symmetric. Rewriting this relationship in matrix form after discretisation, we obtain a solution for the prior q that brings the core component closest to symmetry (Fig 2C). Fig 2D shows the asymmetric part of the resulting core matrix, which clearly contains non-zero residuals near vertical and horizontal. We conclude that there is no prior that would allow posterior sampling to explain these data.

Consistent mappings with changing distributions. In a third experiment we removed the separate positional cue, instead indicating which stimulus was to be recalled by the location of the response grating. The delay between sample and response was randomised. Comparing responses on short delay trials (<2 s) to those on long delay trials (>4 s), we found that response dispersion and bias changed with delay, particularly around vertical (Fig 2E,F). This suggests that $p(X|S)$ changes with delay. However, responses agreed with the row-defined means regardless of delay (Fig 2G). Thus, the change to X over time must be such that the $X \mapsto R$ mapping maintains distributional consistency, suggesting that the representation carries distributional information.

Conclusion

We find that, when performing a continuous orientation recall task, participants report the mean of the likelihood associated with a variable remembered representation, rather than its mode or a random sample. This behaviour is maintained with delay even as the pattern of variability changes, suggesting that the representation itself may carry rich distributional information.

Acknowledgments

The authors would like to thank the Gatsby Charitable Foundation and the Wellcome Trust for funding this work.

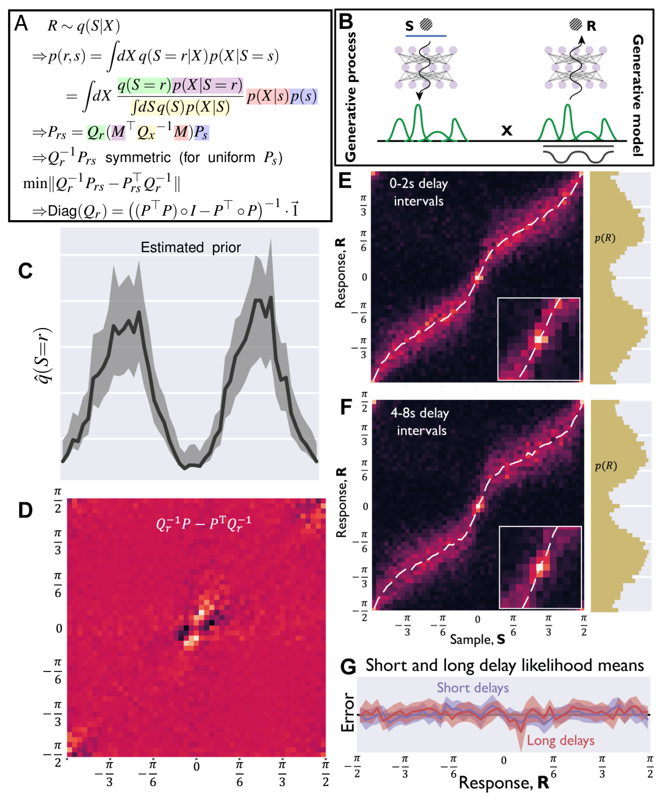

Figure 2: A - Analytic expression for the optimal prior for posterior sampling. Coloured boxes define discretised representations of distributions. B - Schematic of the generative process and model. C - Optimal prior. D - Residuals with the optimal prior. E,F - Joint and response distributions for different delays. G - Mean decision rule errors for different delays.

References

Taylor, R., & Bays, P. M. (2018, July). Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *The Journal of Neuroscience*, *38*(32), 7132–7142. Retrieved from https://doi.org/10.1523/jneurosci.1018-18.2018 doi: 10.1523/jneurosci.1018-18.2018

Wei, X.-X., & Stocker, A. A. (2015, September). A bayesian observer model constrained by efficient coding can explain 'anti-bayesian' percepts. *Nature Neuroscience*, *18*(10), 1509–1517. Retrieved from https://doi.org/10.1038/nn.4105 doi: 10.1038/nn.4105

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., & Stokes, M. G. (2020, March). Drifting codes within a stable coding scheme for working memory. *PLOS Biology*, *18*(3), e3000625. Retrieved from https://doi.org/10.1371/journal.pbio.3000625 doi: 10.1371/journal.pbio.3000625