Humans represent visuo-spatial probability distribution as k-means clusters

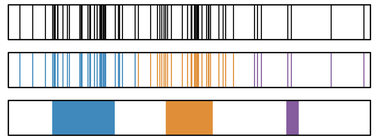
Jingwei Sun[1], Jian Li[1,2], Hang Zhang[1,2,3]

[1] School of Psychological and Cognitive Sciences, Peking University, Beijing, China; [2] PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China;
[3] Peking-Tsinghua Center for Life Science, Peking University, Beijing, China

1 BACKGROUND

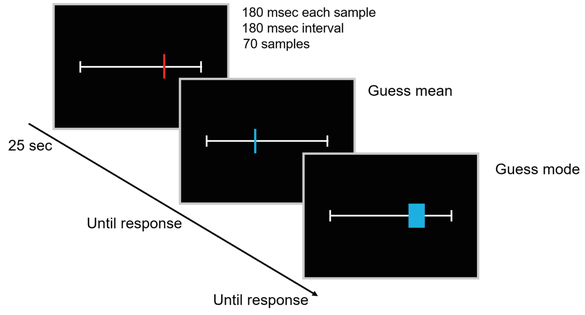
The reward for a real-world choice is often randomly distributed on a continuum (e.g. the future price of a stock). However, little is known about how reward distributions—or, probability distributions in general—are represented in the human brain.[1,2] In this study, we explored potential heuristic representations of probability distributions and tested them on human subjects.

Hypothesis: Humans represent probability distributions as k-means clusters, and estimate their mean and mode according to our proposed **k-Means Sampling Model**.



2 METHODS

Paradigm



Stimuli

Beta-mix distribution: Mixture of several non-overlapping same-shape beta distributions, with 2 sets of parameters: (1) sub-beta distribution parameters (α, β); (2) weights of each sub-beta distribution (w_1, w_2, …, w_n), where $\sum w_i = 1$ (i = 1,2,3,…,n).

Beta(3, 1.5)
Weights = (0.5, 0.3, 0.2)

Experiment 1
(N = 9)
3-beta-mix distribution
(α, β) ∈ {(3.10, 1.05), (2.91, 2.91), (1.05, 3.10)}
Weights: permutation of (0.5, 0.3, 0.2)
18 distribution * 9 samples = 162 trials
Samples presented in random order

Experiment 2
(N = 10)
4-beta-mix distribution
(α, β) ∈ {(3.10, 1.05), (2.91, 2.91), (1.05, 3.10)}
Weights: permutation of (0.4, 0.3, 0.2, 0.1)
72 distribution * 2 samples = 144 trials
Samples presented in random order

3 KEY POINTS OF K-MEANS SAMPLING MODEL

Sampling. Subjects' representation of the spatial distribution of the 70 lines is based on a randomly selected subset of the lines (i.e. random samples). A specific subject's sample size is constant across trials, denoted n_s.

k-means representation. Subjects decompose their samples from the distribution into k non-overlapping clusters (using algorithms close to k-means clustering) and represent the mean and weight of each cluster.

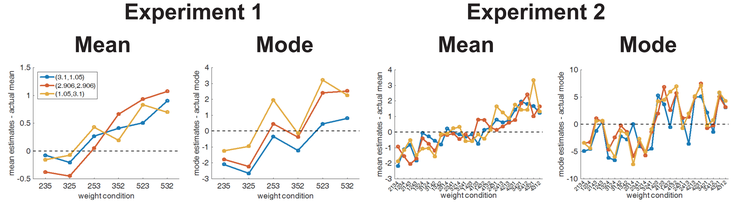
Mode estimates. Subjects report the mean of the cluster with the highest weight. Because of random variations, the reported cluster does not necessarily correspond to the highest mode (among multiple modes). The sample size n_s is estimated for each subject based on their mode estimates.

Mean estimates. Subjects report a weighted average of the means of the clusters. Subjects' estimates may deviate from the objective mean because the subjective weights for the clusters may not agree with the objective weights. The posterior probability of the ordering of subjective weights is calculated for each trial based on n_s and subject's mode estimate.

4 RESULTS

Regression Results

All subjects' mean and mode responses were highly correlated with, but also systematically deviated from the true means and modes.

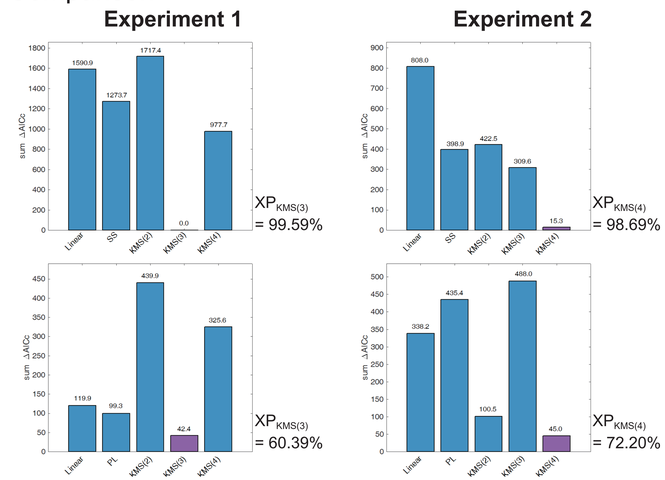


Models Considered

1 Mode estimates
Linear Model. *modeEstimates* ~ $\beta_0 + \beta_1 * actualMean + \beta_2 * actualMode$.
Single sample. Randomly use a single sample as the mode estimate.
k-Means sampling.

2 Mean estimates
Linear Model. *meanEstimates* ~ $\beta_0 + \beta_1 * actualMean + \beta_2 * actualMode$.
Posterior linear. *meanEstimates* ~ $\beta_0 + \beta_1 * actualMean + \beta_2 * modeEstimates$.
k-Means sampling.

Model Comparison



SS: Single Sample; KMS: k-Means Sampling; PL: Posterior Linear; EFR: Error-free Ranking;
XP_{Model}: exceedance probability (probability for the model to be the best model).

k-means sampling model fits best among other competing models. Specifically, subjects are more likely to partition samples into 3 clusters in 3-beta-mix distribution condition, and 4 in 4-beta-mix distribution condition.

5 CONCLUSION

Subjects represent visuo-spatial probability distribution as k-means clusters, and their representation is based on a random subset of the samples.

6 REFERENCES

1 Zhang, H., Daw, N. D., & Maloney, L. T. (2015). Human representation of visuo-motor uncertainty as mixtures of orthogonal basis distributions. *Nature neuroscience*, 18(8), 1152-1158.
2 Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, 16(9), 1170-1178.

Please scan the QR code to assess this poster.

CONTACT: sun.jingwei@pku.edu.cn