Pattern separation using compressed and semantic representations of memory

Dale Zhou (dale.zhou@uci.edu)
University of California, Irvine, Neurobiology and Behavior
519 Biological Sciences Quad, Irvine, 92697, United States

Sharon Noh (nohsm@uci.edu)
Department of Cognitive Sciences
Social Science Lab 334, Irvine, 92697, United States

Michael Yassa (myassa@uci.edu)
University of California, Irvine, Center for the Neurobiology of Learning and Memory
Qureshey Research Laboratory, Irvine, 92697, United States

Aaron Bornstein (aaron.bornstein@uci.edu)
University of California, Irvine, Department of Cognitive Sciences
Social Science Lab 334, Irvine, 92697, United States

Abstract

Recognizing whether information is novel is crucial for many behaviors. For example, decision-makers need to discriminate between perceptually or mnemonically aliased states when planning through uncertain contingencies. In the mnemonic similarity task, images are judged as novel, similar, or old when the presented image was never seen, already seen, or slightly altered as an image "lure." Diminished lure discrimination has been diagnostic of poorer memory in clinical settings, predictive of imprecision in decision-making, and indicates altered pattern separation computations in the hippocampus. However, it is unknown what properties of stimuli drive difficulties in discrimination and whether their influence varies across individuals. We hypothesize that the lure discrimination is related to the image properties of lossy compression, semantic similarity, and intrinsic memorability. Consistent with our hypothesis, we find that participants (n =366) perform better when original and lure images have lossier compression, greater semantic distance, and larger differences in intrinsic memorability. Sensitivity to these image properties tends to worsen with age. Perceptual, semantic, and mnemonic differences may construct distinct memory representations to support pattern separation.

Keywords: memory; efficient coding; lossy compression

Decisions often involve recognizing when information is an example of prior experience or is new. Discriminating among perceptually-aliased observations is a critical step in decision-making, useful for identifying the appropriate current state of the environment in order to select appropriate actions (Bornstein et al., 2023; Khoudary, Peters, & Bornstein, 2022; Noh, Singla, Bennett, & Bornstein, 2023; Aridor, da Silveira, & Woodford, 2024). Discriminating remembered states depends on pattern separation, the ability to distinguish between highly similar inputs with distinct responses (Yassa & Stark, 2011). However, it remains unknown what properties of stimuli characterize pattern separation difficulties across individuals.

Pattern separation can be measured by the mnemonic similarity task (**Figure 1A-B**). Performance is predictive of memory outcomes in clinical settings and the precision of memory representations used for decision-making (Stark, Kirwan, & Stark, 2019; Noh, Cooper, Stark, & Bornstein, 2024). The hippocampus is thought to implement auto-associative and orthogonalizing computations to separate representations as a function of small differences in input (Yassa & Stark, 2011) (**Figure 1C**). This allows for efficiency in stored representations and pattern separation.

Perceptual systems are thought to achieve this efficiency by adapting representations to the properties of inputs, according to the efficient coding hypothesis (Barlow et al., 1961). Due to constraints on information processing, perceptually similar stimuli are lossily compressed into efficient representations. Sims (2018) showed this loss naturally explains when stimuli

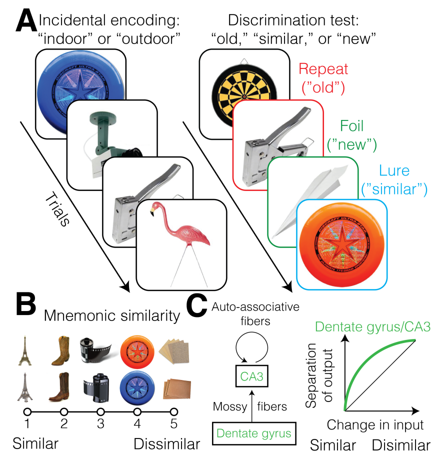

Figure 1: **(A)** Task stimuli with corresponding trial type and correct response. **(B)** Five bins of mnemonic dissimilarity between original and lure images. **(C)** Putative neural function.

are discriminable. Here, we hypothesize that pattern separation can be more richly understood by this and other properties of inputs: their lossy compressibility, semantic relatedness, and intrinsic memorability.

Method

Mnemonic similarity task. Object images are judged as indoor or outdoor during training (**Figure 1A**). Then they are judged as novel, similar, or old, when the presented image was already seen, never seen, or slightly altered ("repeat", "foil", and "lure" trials, respectively). Pattern separation performance is measured as the separation bias: $p(\text{similar}|\text{lure}) - p(\text{similar}|\text{foil})$. Performance in an independent sample was used to discretize the "mnemonic similarity" of original and lure images into 5 lure bins (**Figure 1B**). Participants (n =366; 46 ± 19 years old; 144 men, 218 women) were recruited from Amazon Mechanical Turk.

Lossy compression. We use a convolutional autoencoder to learn compressed memory representations of color images in 256-dimensional latent space (**Figure 2A**). Sims (2018) established that the loss function from efficiently coding (compressing) an input x into an output $\hat{x}$ can explain when two perceptually similar items are confused (generalized) or discriminated across many sensory modalities (**Figure 2B**). We determine the perceptual loss when compressing images $x_{original}$ and x_{lure} in latent space. To simulate the cost of remembering, we linearly interpolate between $x_{original}$ and x_{lure} in latent space. Then, we calculate the KL divergence between pairs of representations, the information theoretic cost of remembering a representation given the optimal encoding of another representation. This produces a matrix for each pair of original and lure images that represents the perceptual information channel, where rows are input representations and columns are output representations. We apply Bayesian inference to

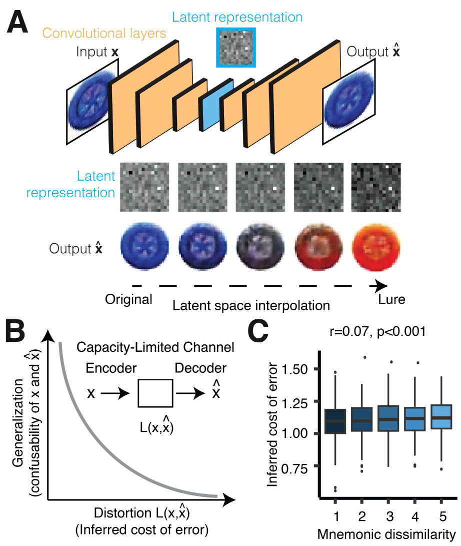

Figure 2: **(A)** Autoencoder's learns compressed latent representations. Discriminability of x and $\hat{x}$ **(B)** theoretically (Sims, 2018) and **(C)** empirically relate to the cost of compression

use the matrix to infer a cost of each matrix input-output element, quantifying the lossiness of compression.

Semantic distance. We use pre-trained transformer models to perform image-to-text conversions, obtaining a 50-dimensional semantic vector embedding per image (Radford et al., 2021). Embeddings are used to calculate the cosine distance between original and lure images (**Figure 3A**).

Intrinsic memorability. Some images have properties that make them easier to remember than others. We used a pre-trained residual network to predict the intrinsic memorability of original and lure images (Needell & Bainbridge, 2022).

Results

Consistent with our hypotheses, original and lure images that are more "distorted" from each other after lossy compression tend to have greater mnemonic dissimilarity (Spearman correlation $\rho = 0.07, p < 0.001$; **Figure 2**). The semantic distance between original and lure images relates to mnemonic dissimilarity ($\rho = 0.29, p < 0.001$; **Figure 3**), while intrinsic memorability does not. Compression loss, semantic distance, and memorability predict pattern separation performance (**Figure 4**). Older adults were less sensitive to changes in each property (steeper slopes; property-by-age interactions $p < 0.001$). Different slopes of semantic distance suggest that older adults can use semantic memory to compensate as their semantic knowledge increases (Park & Reuter-Lorenz, 2009).

Conclusion

The information compression, semantic representation, and intrinsic memorability of images predict pattern separation performance. Modeling these processes may help us better understand decision-making using distinct representations of varying complexity, uncertainty, and content.

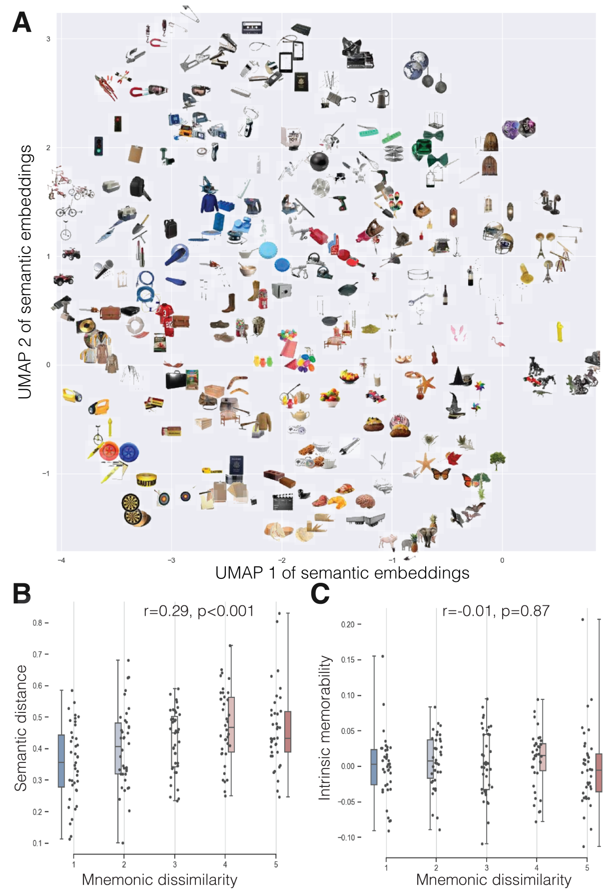

Figure 3: **(A)** Semantic embeddings projected to two dimensions using UMAP. **(B)** Discriminability relates to semantic distance **(C)** but not intrinsic memorability.

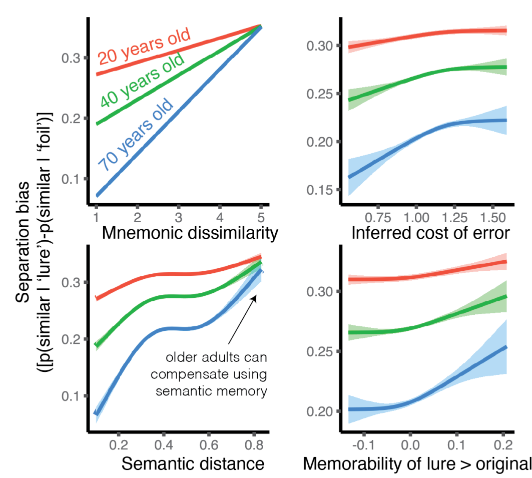

Figure 4: Empirical pattern separation functions with age interactions. Non-linearities were fit using generalized additive models with restricted maximum likelihood. Only the compression function is consistent with the previously hypothesized hippocampal computation (**Figure 1C**). Aging participants depend on starker differences in input, as previously hypothesized (Yassa & Stark, 2011).

Acknowledgments

D.Z. acknowledges funding from the George E. Hewitt Foundation for Medical Research. M.Y. and A.M.B. acknowledge funding from the National Institute for Mental Health (R01MH128306).

References

Aridor, G., da Silveira, R. A., & Woodford, M. (2024). *Information-constrained coordination of economic behavior* (Tech. Rep.). National Bureau of Economic Research.

Barlow, H. B., et al. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, *1*(01), 217–233.

Bornstein, A. M., Aly, M., Feng, S. F., Turk-Browne, N. B., Norman, K. A., & Cohen, J. D. (2023). Associative memory retrieval modulates upcoming perceptual decisions. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(3), 645–665.

Khoudary, A., Peters, M. A., & Bornstein, A. M. (2022). Precision-weighted evidence integration predicts time-varying influence of memory on perceptual decisions. *Cognitive Computational Neuroscience*.

Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior*, *5*(2), 168–184.

Noh, S. M., Cooper, K. W., Stark, C. E., & Bornstein, A. M. (2024). Multi-step inference can be improved across the lifespan with individualized memory interventions. *PsyArXiv*.

Noh, S. M., Singla, U. K., Bennett, I. J., & Bornstein, A. M. (2023). Memory precision and age differentially predict the use of decision-making strategies across the lifespan. *Scientific Reports*, *13*(1), 17014.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual review of psychology*, *60*, 173–196.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... others (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748–8763).

Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science*, *360*(6389), 652–656.

Stark, S. M., Kirwan, C. B., & Stark, C. E. (2019). Mnemonic similarity task: A tool for assessing hippocampal integrity. *Trends in cognitive sciences*, *23*(11), 938–951.

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in neurosciences*, *34*(10), 515–525.