

Why do we have two memory systems? An information-theoretic account

Main text: 722 words

Traditional taxonomies of long-term memory mark a distinction between episodic memory, the capacity to relive events in one’s personal past, and semantic memory, the capacity to remember general facts about oneself or the environment. Recently, such distinction has fallen under heavy scrutiny (Aronowitz 2022; Gentry and Buckner 2024). Given the interactions between episodic and semantic memory under consolidation, and the fact that some memories present both episodic and semantic features, philosophers and cognitive scientists have argued that there is no clean distinction between these memory systems. Instead, these researchers argue that episodic and semantic memory form a continuum. However, what is missing from this literature is an account of *why* this continuum exists and of the computational principles that underlie it. In this paper, I defend that such continuum is adaptive due to the constraints of rate-distortion theory.

Rate-distortion theory is a formal account of the idea that for any representation to be as faithful as possible, it must be as complex as its target. For example, a map of a city will be more accurate the more it encodes details about the city itself. Yet, doing so would diminish its usefulness as a map. As such, there is a trade-off between representational complexity and fidelity. In line with this trade-off, rate-distortion theory asks how many bits of information can one remove without distorting a source beyond a certain point. Where $R \in \mathbb{R}$ is a representation’s rate (roughly equivalent to its length and complexity), and $d : X \times Y \rightarrow \mathbb{R}_{\geq 0}$ is a function that measures how much distortion there is between a representation and its target, Shannon (1959) proves that, for any distortion value d' , there is a single lowest value of R that meets the distortion d' , denoted as $R(d')$. Likewise, for a set rate R , d' is the minimum amount of distortion one can expect for a representation of that complexity.

Rate-distortion theory thus places a hard constraint on information processors. It stipulates that, for any distortion between a representation and its target, the complexity of that representation must be at least $R(d')$. Like other constraints, it sets hard limits on the value of representational complexity. Moreover, the theory is external to how an information processor is implemented, is fixed relative to other factors in the vicinity and guides the explanandum outcome, rather than triggering it. Per Ross (2023), these features are characteristic of hard constraints, which explain why it is impossible for some systems to feature some properties or instantiate some values.

Being systems that process information from the past, episodic and semantic memory have to follow the constraints of rate-distortion theory. Moreover, the continuum between these systems naturally mirrors the range of possible pairs of rate and distortion values. Whereas episodic memories are complex representations (i.e., with higher rates) that faithfully reconstruct the past (hence, with a lower distortion), semantic memories are simple representations (lower rates) that do not capture the details of past events (higher distortion). Having both of these memory systems

is adaptive in so far as they capture both ends of the trade-off: they allow for us to use simple representations when accuracy is not paramount, while having the option to use a more complex representation when we need less distortion. It also allows us to make generalizations and learn patterns without the cost of losing specific information about past events.

Moreover, one consequence of rate-distortion theory is that overlapping representations are more efficiently encoded as one representation (Berger 1971). Some experiments suggest that our memory systems function in this way. The Mnemonic Similarity Test was designed specifically to measure participants' ability in recognizing old stimuli and telling them apart from similar, though different, items (Kirwan and Stark 2007). Different studies in this paradigm have consistently shown that target repetition in the encoding phase is correlated with participants' misclassification of lures as old items (Reagh and Yassa 2014). This is predicted from rate-distortion theory: correlated and overlapping messages can be lumped into a single representation for efficient encoding. While this happens at the cost of higher distortion, it is a way for our memory system to spare the amount of resources dedicated to remembering the past.

In summary, rate-distortion theory provides a formal account of why episodic and semantic memory systems form a continuum. According to this account, such continuum is adaptive because it allows for our cognitive system to flexibly use representations that fall on either end of the rate-distortion curve.

References

- Aronowitz, Sara (July 19, 2022). "Semanticization Challenges the Episodic–Semantic Distinction". In: *The British Journal for the Philosophy of Science*, p. 721760. ISSN: 0007-0882, 1464-3537. DOI: 10.1086/721760. URL: <https://www.journals.uchicago.edu/doi/10.1086/721760> (visited on 01/18/2023).
- Berger, Toby (1971). *Rate Distortion Theory; a Mathematical Basis for Data Compression*. Englewood Cliffs, N.J., Prentice-Hall. 346 pp. ISBN: 978-0-13-753103-5. URL: <http://archive.org/details/ratedistortionh0000berg> (visited on 11/14/2025).
- Gentry, Hunter and Cameron Buckner (Sept. 16, 2024). "Transitional Gradation and the Distinction between Episodic and Semantic Memory". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379.1913, p. 20230407. DOI: 10.1098/rstb.2023.0407. URL: <https://royalsocietypublishing.org/doi/full/10.1098/rstb.2023.0407> (visited on 08/19/2025).
- Kirwan, C. Brock and Craig E. L. Stark (Jan. 9, 2007). "Overcoming Interference: An fMRI Investigation of Pattern Separation in the Medial Temporal Lobe". In: *Learning & Memory* 14.9, pp. 625–633. ISSN: 1072-0502-01, 1549-5485. DOI: 10.1101/lm.663507. PMID: 17848502. URL: <http://learnmem.cshlp.org/content/14/9/625> (visited on 09/05/2025).
- Reagh, Zachariah M. and Michael A. Yassa (Jan. 7, 2014). "Repetition Strengthens Target Recognition but Impairs Similar Lure Discrimination: Evidence for Trace Competition". In: *Learning & Memory* 21.7, pp. 342–346. ISSN: 1072-0502, 1549-5485. DOI: 10.1101/lm.034546.114. PMID: 24934334. URL: <http://learnmem.cshlp.org/content/21/7/342> (visited on 01/26/2024).
- Ross, Lauren N. (Aug. 10, 2023). "The Explanatory Nature of Constraints: Law-based, Mathematical, and Causal". In: *Synthese* 202.2, p. 56. ISSN: 1573-0964. DOI: 10.1007/s11229-023-04281-5. URL: <https://doi.org/10.1007/s11229-023-04281-5> (visited on 04/10/2025).
- Shannon, C. E. (1959). "Coding Theorems for a Discrete Source with a Fidelity Criterion". In: *IRE National Convention Record* 7, pp. 325–350.