Effect of Retention Interval on Mnemonic Processing in the Mnemonic Similarity Task

Atharva Sunil Gogate (atharva.gogate@research.iiit.ac.in),
Hurshitha Vasudevan (hurshitha.vasudevan@research.iiit.ac.in) &
Bhaktee Dongaonkar (bhaktee.dongaonkar@iiit.ac.in)
Cognitive Science Lab, International Institute of Information Technology, Hyderabad

Abstract

Pattern separation enables discrimination between similar experiences and relies on hippocampal function. The Mnemonic Similarity Task (MST) indexes this process using the Lure Discrimination (LDI). Recent studies, on the other hand, suggest MST performance is influenced by perceptual processing. We investigated the extent of mnemonic processes in MST performance across increasing delays between encoding and retention phases (0 minutes, 30 minutes, 8 hours, 24 hours). LDI declined in the 8-hour and 24-hour groups, with significantly worse lure discrimination especially for low similarity lures. Response patterns revealed a bias, shorter delays showed an "Old" response bias, whereas longer delays were characterized by a "New" response bias. Additionally, decreasing lure similarity was associated with increased "New" responses. Together, these findings suggest that the MST is sensitive to mnemonic processing over short delays, however performance at longer delays is constrained and may reflect non-mnemonic factors.

Keywords: Pattern separation; Hippocampus; Mnemonic Similarity Task (MST); Lure Discrimination Index (LDI); Mnemonic processing; Delay-dependent memory; Recognition Score (REC)

Introduction

Pattern separation is the process of distinguishing similar experiences and forming distinct representations for storage and retrieval. Overlapping yet distinct representations are formed by the dentate gyrus and CA3 sub-regions of the hippocampus. The hippocampus is susceptible to aging and neurodegenerative conditions. Thus, assessing pattern separation abilities is important to detect changes in the underlying hippocampal function.

The Mnemonic Similarity Task (MST) (Davidson et al., 2018; Jensen et al., 2023; Pishdadian et al., 2020; Stark et al., 2015, 2019; Yassa et al., 2011) is a widely used task designed to assess pattern separation and completion. In the MST, participants distinguish between previously seen items (targets), visually similar items (lures), and novel items (foils). Successful pattern separation entails identifying lure items not as "Old", which is considered a misclassification, but as "Similar". This ability is typically quantified via the Lure Discrimination Index (LDI), calculated by subtracting the probability of misidentifying foils as "Similar" from the probability of correctly identifying lures as "Similar" ($LDI = p(\text{"}Similar\text{"}|Lure) - p(\text{"}Similar\text{"}|Foil)$). Successful lure discrimination requires high-fidelity mnemonic representations, and is primarily thought to be aided by the dentate gyrus and the CA3 sub-regions (Doxey et al., 2019; Hanert et al., 2017; Pishdadian et al., 2020; Stark et al., 2019; Yassa et al., 2011). On the other hand, the recognition memory is calculated via the Corrected Recognition Score (REC) as the probability of identifying a target as "Old" minus the probability of misidentifying a foil as "Old" ($REC = p(\text{"}Old\text{"}|Target) - p(\text{"}Old\text{"}|Foil)$). Recognition memory serves as the proxy for pattern completion, which is the process by which the brain retrieves a stored representation from a partial or incomplete perceptual cue. It is also referred as familiarity-based memory.

Previous reports suggest that LDI scores were significantly lower in older adults compared to younger adults. Age-related changes in the medial temporal lobe and declining hippocampal function in older adults were shown to negatively affect MST performance (Stark et al., 2015, 2019; Yassa et al., 2011).

Recent studies have reported that the LDI score correlated with the visual perception composite score (Davidson et al., 2018) in older adults. Another study used the Montreal Cognitive Assessment (MoCA) along with the MST in an aging population and found that LDI did not correlate to the memory subscale of the MoCA, thus challenging the specificity of the MST with respect to hippocampal based memory (Pishdadian et al., 2020). With only correlation studies, it is unclear to what extent performance in the MST reflects perceptual processing versus mnemonic processing.

Several studies involving the MST investigated age-related decline and the task's senstivity to mnemonic processing is established across various populations. Hanert et al. (2017) performed a variant of the MST on university students with a 9-hour gap between the encoding and recall phases involving either nocturnal sleep or daytime wakefulnes. They found that sleep-based consolidation reduced deterioration of LDI for low-similarity lures over the wakefulness condition. However, this study had a low sample size and the recall phase was conducted twice: immediately after the encoding phase and then after the delay period, possibly leading to practice effect. Doxey et al. (2019) tested young adults with a similar experimental design to that of Hanert et al. (2017), opting for a 12-hour window between encoding and recall, also with similar results, i.e., sleep-based retention preserved pattern separation better than a wakeful delay period.

Mnemonic information is made available for future retrieval by forming a distinct representation. In the typical MST, sev-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

eral images of objects without a background or context are presented briefly, one at a time, followed by a recognition task. The structure of the task may result in insufficient encoding with poor mnemonic representations, with more reliance on perceptual processing. Immediate lure discrimination may remain intact, as previous studies have shown (Stark et al., 2015; Yassa et al., 2011) due to perceptual processing (Lech & Suchan, 2014), but delayed lure discrimination may be affected by poor mnemonic representations and perceptual decay. To confirm the extent of mnemonic processing in MST, we tested the recognition phase of MST at different intervals after the encoding phase. If the MST is a mnemonic task, then the mnemonic effects must persist at longer delays. This study introduced 4 different delay intervals (0 minutes, 30 minutes, 8 hours, 24 hours) between the encoding and recognition phases. We hypothesized that lure discrimination would be unaffected at the immediate and 30-minute delays. At longer delays (8 hours and 24 hours), lure discrimination would be lower due to poor mnemonic representations from insufficient encoding. We also hypothesized that increasing the retention interval would selectively reduce lure discrimination for high-similarity lures, with minimal effects on low-similarity lures and recognition accuracy. Low similarity lures can be discriminated with coarse representations, but high-similarity lures require well-defined representations which may be difficult to form given the insufficient encoding.

Methods

Participants

This study was approved by the Institutional Ethics Committee. 120 young adults (age range: 18-27 years old) participated in this study, all of whom were Indian residents and students of the International Institute of Information Technology, Hyderabad. Participants were screened for mental health, previous night's sleep quality, and alcohol, caffeine, and nicotine consumption before the task.

Mnemonic Similarity Task

We used the MST outlined by Stark et al. (2019).

Experiment Structure

1. Encoding Phase: Participants viewed 128 images of common objects on a white background and classified them as "indoor" or "outdoor".
2. Delay Phase: A retention interval of 0 minutes, 30 minutes, 8 hours, or 24 hours was introduced between the encoding and recall phases.
3. Recall Phase: Participants were shown 64 repetitions/targets, 64 lures, and 64 foils in a randomized order, which they had to classify as "Old", "Similar", or "New", respectively.

In both phases, stimuli were displayed on the screen for 2000 ms, followed by a 500-ms inter-stimulus interval.

Participants were randomly assigned to one of the four delay groups, whose summary is provided in the following table:

Table 1: Distribution of participants into delay groups. "Delay" refers to the retention interval introduced between the encoding and recall phases.

Delay group	Delay	*N*	Men/Women	Mean age + SD
Immediate	0 min	30	5M / 25W	22.30 ± 2.94
Del_30min	30 min	30	16M / 14W	22.73 ± 2.65
Del_8hr	8 hr	30	23M / 7W	20.20 ± 1.77
Del_24hr	24 hr	30	14M / 16W	23.80 ± 2.31

Following is the rationale for selecting the four retention intervals:

1. **0 minutes** (*Immediate*): This served as the baseline for immediate mnemonic and perceptual sensitivity and for potential comparison against other studies utilizing the MST. Peak performance was expected due to minimal temporal interference and active short-term/iconic memory.
2. **30 minutes** (*Del_30min*): This was chosen to test the initial preservation of representations. It allowed for the decay of iconic memory while assessing if the high-fidelity features of the stimuli were preserved on the short term.
3. **8 hours** and **24 hours** (*Del_8hr*, *Del_24hr*): These extended delays highlighted the limit of the MST's design for long-term consolidation. Introducing such retention intervals allowed us to investigate whether the drop in MST performance was due to mnemonic decay of the trace or if the initial encoding was insufficient to maintain long-term retention.

Lure Similarity Bins The lure stimuli were placed into 5 equally sized lure bins (L1-L5) based on their similarity to their respective originally encoded item (Stark et al., 2019; Yassa et al., 2011) Lures became progressively more dissimilar from bins L1 to L5. L1 and L2 represented lures that were relatively more similar to encoding-phase objects, whereas L4 and L5 represented lures that were less similar to the encoding-phase objects.

Results

The Lure Discrimination Index (LDI) scores for each delay group were calculated as $p(\text{"}Similar\text{"}|Lure) - p(\text{"}Similar\text{"}|Foil)$. The subtracted term is a bias correction metric to account for potential "Similar" response bias, as it provides a neutral baseline (Stark et al., 2019).

To test differences in LDI across delay intervals, a one-way ANOVA was conducted with the delay intervals as the independent variable and LDI as a dependent variable (Figure 1). We found a significant main effect of delay intervals ($F(3, 116) = 8.93, p < 0.001, \eta^2 = 0.19$). Follow-up Tukey HSD tests indicated that *Immediate*, *Del_30min*, and *Del_8hr* exhibited significantly higher LDI scores compared to *Del_24hr* ($p < 0.001, p < 0.001, p < 0.01$, respectively). Additionally, LDI scores of *Immediate*, *Del_30min*, and *Del_8hr* were not significantly different from each other ($p > 0.1$ for all relevant comparisons).

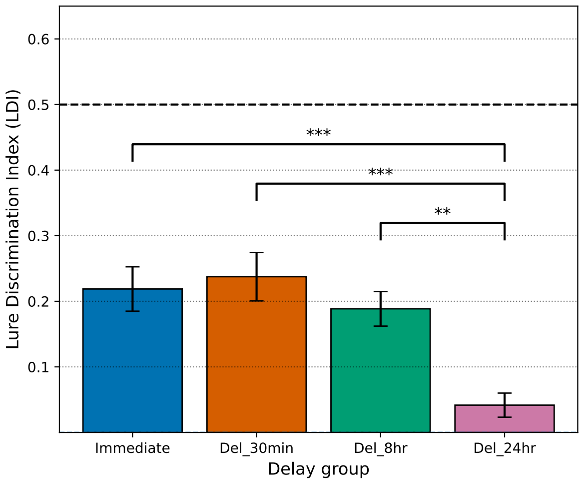

Figure 1: Lure Discrimination Index (LDI) across delay groups. Error bars denote standard error of the mean (SEM)

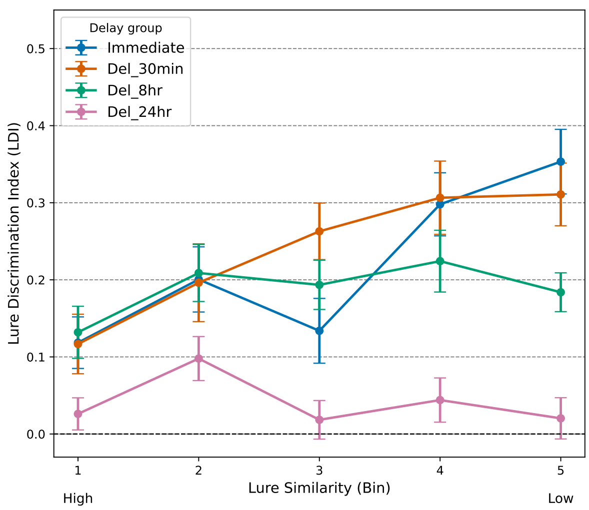

Figure 2: Lure Discrimination Index (LDI) across delay groups and lure bins. Error bars denote standard error of the mean (SEM)

To assess the effect of delay intervals on LDI performance across lure bins, a multivariate ANOVA was performed (Figure 2), indicating a significant main effect ($F(15,342) = 5.40, p < 0.001, \eta^2 = 0.19$). Follow-up univariate ANOVAs showed significant effects of delay intervals in L3 ($F(3,116) = 9.02, p < 0.001, \eta^2 = 0.19$), L4 ($F(3,116) = 9.33, p < 0.001, \eta^2 = 0.19$), and L5 ($F(3,116) = 18.81, p < 0.001, \eta^2 = 0.33$). Post-hoc Tukey HSD comparisons (Table 2) indicated that in L3, *Del_30min* and *Del_8hr* performed significantly better than *Del_24hr* (both $p < 0.001$). In L4 and L5, *Immediate*, *Del_30min*, and *Del_8hr* had significantly better performance than *Del_24hr* (all $p < 0.001$). Additionally, *Immediate* showed significantly higher LDI than *Del_8hr* in L5 ($p < 0.01$).

Table 2: Significant Tukey HSD post-hoc comparisons for LDI across lure bins shown in Figure 2. Group 2 always reflects higher mean LDI than Group 1.

Lure bin	Group 1	Group 2	Significance
Bin 3	*Immediate*	*Del_30min*	$p < 0.05^{*}$
Bin 3	*Del_24hr*	*Del_30min*	$p < 0.001^{***}$
Bin 3	*Del_24hr*	*Del_8hr*	$p < 0.01^{**}$
Bin 4	*Del_24hr*	*Immediate*	$p < 0.001^{***}$
Bin 4	*Del_24hr*	*Del_30min*	$p < 0.001^{***}$
Bin 4	*Del_24hr*	*Del_8hr*	$p < 0.01^{**}$
Bin 5	*Del_8hr*	*Immediate*	$p < 0.01^{**}$
Bin 5	*Del_24hr*	*Immediate*	$p < 0.001^{***}$
Bin 5	*Del_24hr*	*Del_30min*	$p < 0.001^{***}$
Bin 5	*Del_24hr*	*Del_8hr*	$p < 0.01^{**}$

The Corrected Recognition Scores (REC) for each delay group were calculated as $p(\text{"}Old\text{"}|Target) - p(\text{"}Old\text{"}|Foil)$. Similar to the LDI, the subtracted term is a bias correction metric to account for potential "Old" response bias, providing a neutral baseline. To compare the REC across delay groups (Figure 3), a one-way ANOVA was conducted, revealing a significant main effect of group ($F(3,116) = 39.92, p < 0.001, \eta^2 = 0.56$). Post-hoc Tukey HSD tests revealed that both *Immediate* and *Del_30min* performed significantly better than both *Del_8hr* and *Del_24hr* ($p < 0.001$ for all relevant comparisons). *Immediate* and *Del_30min* had comparable REC. Similarly, *Del_8hr* and *Del_24hr* did not significantly differ from each other. REC and LDI scores did not correlate at any delay interval.

To examine whether overall response performance differed across delay intervals, a multivariate ANOVA was conducted with "Old", "Similar", "New" responses as dependent variables and delay groups as the between-subjects factor. This analysis revealed a significant effect of delay group on MST retention ($F(9,348) = 10.60, p < 0.001, \eta^2 = 0.22$). Follow-up univariate ANOVAs indicated significant differences across delay groups; "Old" responses ($F(3,116) = 16.68, p < 0.001, \eta^2 = 0.30$), "Similar" responses ($F(3,116) = 4.98, p = 0.003, \eta^2 = 0.11$), and "New" responses ($F(3,116) = 28.98, p < 0.001, \eta^2 = 0.43$). Post-hoc Tukey HSD tests (Table 3) showed that *Immediate* had significantly higher "Old" responses than *Del_8hr* and *Del_24hr* ($p < 0.001, p < 0.001$, respectively), and *Del_30min* had significantly more "Old" responses than *Del_24hr* ($p < 0.001$). On the other hand, *Del_8hr* and *Del_24hr* exhibited significantly higher "New" responses than both *Immediate* and *Del_30min* ($p < 0.001$ for all relevant comparisons).

The patterns for all responses ("Old", "Similar", "New") across lure bins for delay intervals were studied using Generalized Estimating Equation (GEE) analyzes (Figure 5), which showed that *Del_8hr* and *Del_24hr* when compared to

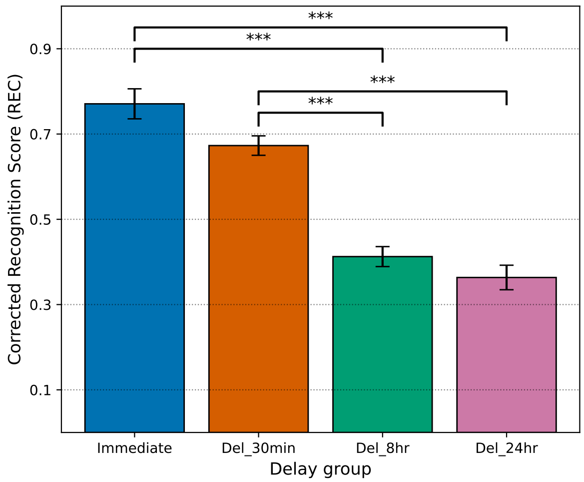

Figure 3: Corrected Recognition Score (REC) across delay groups. Error bars denote standard error of the mean (SEM)

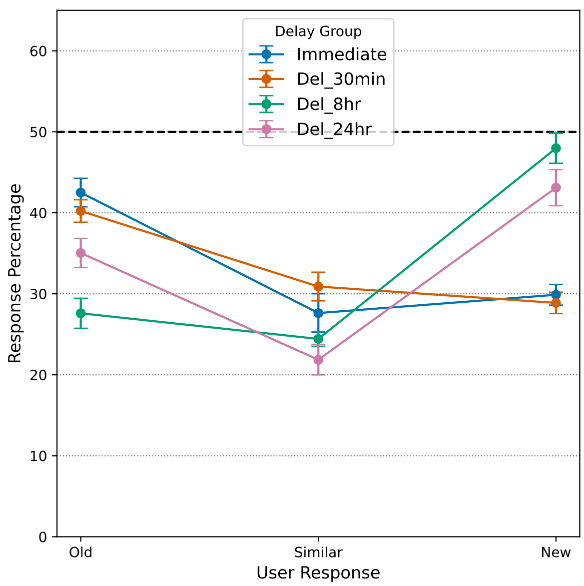

Figure 4: Response distributions to stimuli across delay groups. Error bars denote standard error of the mean (SEM)

Immediate (reference group) had significantly fewer "Old" responses (*Del_8hr*: $\beta = -1.15, p < 0.001$, *Del_24hr*: $\beta = -1.02, p < 0.011$) and higher "New" responses (*Del_8hr*:$\beta = 1.85, p < 0.001$, *Del_24hr*:$\beta = 1.85, p < 0.001$). "Similar" responses significantly reduced in *Del_8hr* ($\beta = -0.176, p < 0.001$)and *Del_24hr* ($\beta = -0.266, p < 0.001$).

Lure similarity had a positive main effect ($\beta = 0.231, p < 0.001$). From L1 to L5, "New" responses increased ($\beta = 0.244, p < 0.001$). However, at *Del_8hr* ($\beta = -0.12, p < 0.01$) and *Del_24hr* ($\beta = -0.11, p < 0.05$), significant negative interactions were observed which indicated a generalized "New" bias for lure stimuli, irrespective of lure bins.

Table 3: Significant Tukey HSD post-hoc comparisons for response proportions across delay groups shown in Figure 4. Group 2 always reflects a higher mean than Group 1.

Response	Group 1	Group 2	Significance
Old	*Del_8hr*	*Immediate*	$p < 0.001^{***}$
Old	*Del_24hr*	*Immediate*	$p < 0.01^{**}$
Old	*Del_8hr*	*Del_30min*	$p < 0.001^{***}$
Old	*Del_8hr*	*Del_24hr*	$p < 0.05^{*}$
Similar	*Del_24hr*	*Del_30min*	$p < 0.01^{**}$
New	*Immediate*	*Del_8hr*	$p < 0.001^{***}$
New	*Immediate*	*Del_24hr*	$p < 0.001^{***}$
New	*Del_30min*	*Del_8hr*	$p < 0.001^{***}$
New	*Del_30min*	*Del_24hr*	$p < 0.001^{***}$

Discussion

This study aimed to understand the extent of mnemonic processing in the MST. We hypothesized that the MST does not support sufficient encoding which would result in poor lure discrimination at longer retention intervals, at shorter intervals the lure discrimination may stay intact due to perceptual processing benefits. This was confirmed by our results, which showed significantly worse lure discrimination in *Del_24hr* compared to *Immediate*, *Del_30min*, and *Del_8hr*. *Immediate*, *Del_30min*, and *Del_8hr* did not show significant differences from each other (Figure 1).

Next, we hypothesized that at longer delay intervals, high similarity lure discrimination would be most affected while low similarity lures discrimination would remain unaffected. Instead we observed a surprising trend that directly contrasted the results observed by Hanert et al. (2017) where LDI scores of highly dissimilar lures following a 9-hour retention interval involving nocturnal sleep or daytime wakefulness were significantly better after sleep. Their result comes with the caveat that all participants completed both variations of the MST, likely leading to practice effect during recall after sleep. In our study, across all lures, a floor effect was observed in LDI scores of the longer delay groups (Figure 2). Participants faced difficulty in discriminating even the low similarity lures L4 and L5 (Figure 2). This supported the hypothesis that insufficient encoding resulted in impoverished representations across all lure bins. "Similar" responses also significantly reduced to below chance across all lure bins at *Del_24hr* (Figure 5) suggesting dependence on perceptual processes.

To test the presence of bias-driven patterns over mnemonic strategies at longer retention intervals, additional analysis was conducted which confirmed that participants in the longer delay intervals (*Del_8hr* and *Del_24hr*) consistently had a higher proportion of "New" responses ("New" bias); whereas the shorter interval groups (*Immediate* and *Del_30min*) consistently had a higher proportion of "Old" responses ("Old" bias) (Figure 4). This aligns with the stance of perceptual dependence for lure discrimination.

There was an increased tendency for the longer delay groups to respond "New" to targets and lures, implying an encod-

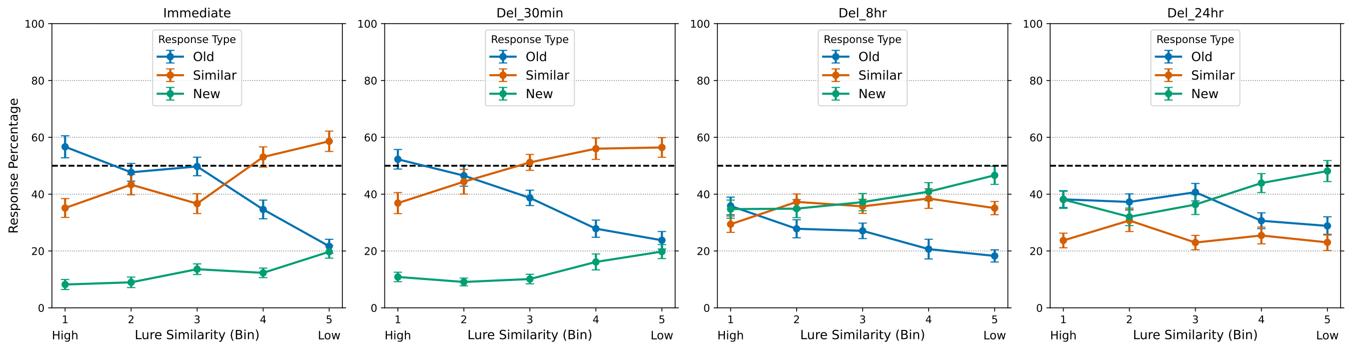

Figure 5: Response distribution to lures across delay groups and lure bins. This graph is indicative of the trends found in the GEE analysis. Error bars denote standard error of the mean (SEM)

ing issue. This was further strengthened by the fact that REC sharply declined for the longer delay groups (*Del_8hr* and *Del_24hr*) (Figure 3). These findings are consistent with Davidson et al. (2018), which showed worse recognition scores for younger adults and showed that MST performance was positively correlated with the visual perception composite rather than long-term episodic memory or executive function.

The intermediate performance of *Del_8hr* (reduced LDI scores relative to the shorter delay groups but significantly higher than *Del_24hr*) indicated partial preservation of mnemonic information. Simultaneously, response distributions resembled those of *Del_24hr* more closely than those of *Del_30min*, suggesting a shift in response strategy. Thus, *Del_8hr* seems to reflect a transition phase from short-term mnemonic strategies to peceptual and bias-driven strategies. This hints at a lack of specificity of the MST with respect to memory (Pishdadian et al., 2020), and more of the lingering traces visual perceptual memory (Davidson et al., 2018).

This interpretation is supported by the lure bin analyses, where GEE models revealed that increasing the retention interval significantly reduced the likelihood of "Old" responses while increasing the probability of "New" responses, independent of lure similarity. Shorter intervals had higher "Similar" responses for high similarity lures, while this behavior attenuated for longer intervals. This suggests that longer intervals led to a reliance on novelty judgments over mnemonic discrimination.

These results challenge the assumption that MST is a mnemonic task primarily dependent on hippocampal function. While LDI scores in *Del_8hr* were not significantly worse than those in the shorter retention intervals, the lure discrimination trend (Figure 5) was similar to that of *Del_24hr*. The novelty bias and the poor LDI scores at longer intervals suggest lingering perceptual discrimination (Davidson et al., 2018) in higher visual processing areas (Lech & Suchan, 2014), instead of hippocampal dependent pattern separation. Stimuli that are not feature-heavy can be processed and retained for brief periods by the higher visual processing areas. When stimuli are highly similar and overlapping, the demands on pattern separation increase and hippocampal processing may be required (Lech & Suchan, 2014).

Overall, our findings suggest that the design of the MST has limitations. Participants viewed a large number of images for 2 seconds each, devoid of context, which could favor regions of higher visual processing capable of short-term discrimination without hippocampal encoding (Lech & Suchan, 2014). Performance at shorter retention intervals reflect perceptual processes and limited involvement of the hippocampal function, and larger retention intervals reveal the limits of such processing. Consistent with previous studies which showed that MST performance was not specific to the memory component (Pishdadian et al., 2020), instead correlated with the visual perception composite (Davidson et al., 2018), the results suggest that the standard MST captures short-term mnemonic sensitivity, not stable long-term pattern separation. The findings in this study additionally suggest a multifactorial model of memory retrieval. Instead of gauging MST performance as solely reflective of hippocampal function, it would be worthwhile to consider a memory model where successful pattern separation or completion is distributed across perceptual, executive, and mnemonic memory processes. While the task remains informative of broader cognitive decline, our findings suggest that MST performance alone may not provide a specific index of hippocampal-dependent memory.

Acknowledgements

We acknowledge the utility of AI tools in the preparation of this manuscript, specifically to improve the coherence of the text. All analyses, results, interpretations, and conclusions are the sole responsibility of the authors.

References

Davidson, P. S. R., Vidjen, P., Trincao-Batra, S., & Collin, C. A. (2018). Older adults' lure discrimination difficulties on the mnemonic similarity task are significantly correlated with their visual perception. *The Journals of Gerontology: Series B*, *74*, 1298–1307. https://doi.org/10.1093/geronb/gby130

Doxey, C. R., Hodges, C. B., Bodily, T. A., Muncy, N. M., & Brock Kirwan, C. (2019). The effects of sleep on the neural correlates of pattern separation [<i>hippocampus</i> , <i>28</i> (2), 108–120 (2018)]. *Hippocampus*, *29*(2), 141–142. https://doi.org/10.1002/hipo.23038

Hanert, A., Weber, F. D., Pedersen, A., Born, J., & Bartsch, T. (2017). Sleep in humans stabilizes pattern separation performance. *The Journal of Neuroscience*, *37*, 12238–12246. https://doi.org/10.1523/jneurosci.1189-17.2017

Jensen, A., Karpov, G., Collin, C. A., & Davidson, P. S. R. (2023). Executive function predicts older adults' lure discrimination difficulties on the mnemonic similarity task (A. Krendl, Ed.). *The Journals of Gerontology: Series B*, *78*(10), 1642–1650. https://doi.org/10.1093/geronb/gbad091

Lech, R. K., & Suchan, B. (2014). Involvement of the human medial temporal lobe in a visual discrimination task. *Behavioural Brain Research*, *268*, 22–30. https://doi.org/10.1016/j.bbr.2014.03.030

Pishdadian, S., Hoang, N., Baker, S., Moscovitch, M., & Rosenbaum, R. (2020). Not only memory: Investigating the sensitivity and specificity of the mnemonic similarity task in older adults. *Neuropsychologia*, *149*, 107670. https://doi.org/10.1016/j.neuropsychologia.2020.107670

Stark, S. M., Kirwan, C. B., & Stark, C. E. (2019). Mnemonic similarity task: A tool for assessing hippocampal integrity. *Trends in Cognitive Sciences*, *23*, 938–951. https://doi.org/10.1016/j.tics.2019.08.003

Stark, S. M., Stevenson, R., Wu, C., Rutledge, S., & Stark, C. E. L. (2015). Stability of age-related deficits in the mnemonic similarity task across task variations. *Behavioral Neuroscience*, *129*, 257–268. https://doi.org/10.1037/bne0000055

Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. (2011). Pattern separation deficits associated with increased hippocampal ca3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*, 968–979. https://doi.org/10.1002/hipo.20808