Received: 12 July 2018 | Revised: 15 August 2018 | Accepted: 13 September 2018

DOI: 10.1002/hipo.23030

WILEY

RESEARCH ARTICLE

Memory Image Completion: Establishing a task to behaviorally assess pattern completion in humans

Paula Vieweg[1,2] | Martin Riemer[2,3,4] | David Berron[2,5,6] | Thomas Wolbers[2,3,4]

[1]Institute of Psychology, University of Leipzig, Leipzig, Germany

[2]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany

[3]Medical Faculty, University Hospital Magdeburg (FME), Otto von Guericke University Magdeburg, Magdeburg, Germany

[4]Center for Behavioral Brain Sciences, Magdeburg, Germany

[5]Institute of Cognitive Neurology and Dementia Research, Otto von Guericke University Magdeburg, Magdeburg, Germany

[6]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden

Correspondence
Paula Vieweg, Institute of Psychology, University of Leipzig, Neumarkt 9–19, Leipzig 04109, Germany.
Email: paula.vieweg@uni-leipzig.de

Funding information
European Research Council Starting Grant, Grant/Award Number: 335090; European Social Fonds (ESF) (Sachsen-Anhalt Wissenschaft Spitzenforschung/Synergien: AGETIME)

Abstract

For memory retrieval, pattern completion is a crucial process that restores memories from partial or degraded cues. Neurocognitive aging models suggest that the aged memory system is biased toward pattern completion, resulting in a behavioral preference for retrieval over encoding of memories. Here, we built on our previously developed behavioral recognition memory paradigm—the Memory Image Completion (MIC) task—a task to specifically target pattern completion. First, we used the original design with concurrent eye-tracking in order to rule out perceptual confounds that could interact with recognition performance. Second, we developed parallel versions of the task to accommodate test settings in clinical environments or longitudinal studies. The results show that older adults have a deficit in pattern completion ability with a concurrent bias toward pattern completion. Importantly, eye-tracking data during encoding could not account for age-related performance differences. At retrieval, spatial viewing patterns for both age groups were more driven by stimulus identity than by response choice, but compared to young adults, older adults' fixation patterns overlapped more between stimuli that they (wrongly) thought had the same identity. This supports the observation that older adults choose responses perceived as similar to a learned stimulus, indicating a bias toward pattern completion. Additionally, two shorter versions of the task yielded comparable results, and no general learning effects were observed for repeated testing. Together, we present evidence that the MIC is a reliable behavioral task that targets pattern completion, that is easily and repeatedly applicable, and that is made freely available online.

KEYWORDS

cognitive aging, eye-tracking, pattern completion bias, recognition memory

1 | INTRODUCTION

In everyday life, we constantly recognize people, places, and objects. Often, the cues that trigger such retrieval are incomplete or different from their original encounter. For example, imagine running into an old friend you have not seen since school. Despite her matured appearance, different style and out-of-school context, recognizing her and resurfacing old memories are seamlessly enabled by a process called pattern completion.

Pattern completion and separation have received widespread attention in neuroscience to foster the understanding of how the brain encodes and retrieves memories (for extensive reviews see Hunsaker & Kesner, 2013; Yassa & Stark, 2011). Based on computational models (Hasselmo & McClelland, 1999; Marr, 1971; McClelland, 1994; O'Reilly, Norman, & McClelland, 1998; Treves & Rolls, 1994) and findings in rodent studies (Guzowski, Knierim, & Moser, 2004; Leutgeb & Leutgeb, 2007; Leutgeb, Leutgeb, Moser, & Moser, 2007; Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Nakazawa et al., 2002; Neunuebel & Knierim, 2014; Vazdarjanova & Guzowski, 2004), pattern completion is thought to restore memory traces from partial or degraded input via neural reactivation, mainly originating in hippocampal region CA3. In contrast, pattern separation is considered as the complimentary process of differentiating very similar memories via orthogonalization of neural representations, which predominantly occurs in the dentate gyrus (DG).

While both processes are defined based on neural computations, their behavioral outcomes have been approximated in human research with behavioral tasks. A central paradigm is the mnemonic similarity task (MST; Stark, Stevenson, Wu, Rutledge, & Stark, 2015), in which participants are presented with very similar stimuli (lures) among novel and repeated stimuli (Bakker, Kirwan, Miller, & Stark, 2008; Kirwan & Stark, 2007; Lacy, Yassa, Stark, Muftuler, & Stark, 2011; etc.). While the MST efficiently induces memory interference, the use of highly similar stimuli during encoding selectively favors pattern separation because similar memory traces need to be differentiated. It is possible that pattern completion is merely a secondary finding in these studies, that is, without being explicitly manipulated. Hunsaker and Kesner (2013 p. 40) have indeed suggested that the presentation of parts of an original cue may engage pattern completion more independently than the presentation of similar cues. Additionally, they suggested that even though both processes likely contribute to memory encoding and retrieval, pattern separation may be more involved in the former and pattern completion in the latter. Thus, in order to increase the impact of either pattern separation or completion, studies should focus on the respective corresponding phase of memory processing. This idea received additional support from an eye-tracking study (Molitor, Ko, Hussey, & B a, 2014), showing that false alarm trials in the MST (i.e., a similar stimulus judged as old), which were interpreted to involve pattern completion, were associated with fewer fixations during encoding. This observation suggests an additional involvement of pattern separation processes.

Accommodating the above suggestions, we have previously developed the Memory Image Completion (MIC) task (Vieweg, Stangl, Howard, & Wolbers, 2015; named MIC in Baker et al., 2016). In this paradigm, gradually less complete versions of a learned visual stimulus reduce recognition performance, which was interpreted to reflect an increase in pattern completion demands. The practical suitability of our task was later independently confirmed in a comparative meta-analysis (Liu, Gould, Coulson, Ward, & Howard, 2016), and it has also been employed successfully in a lesion study (Baker et al., 2016). Note that, by now, there are some other paradigms tapping into pattern completion (Horner & Burgess, 2014; Paleja & Spaniol, 2013; Staresina, Alink, Kriegeskorte, & Henson, 2013). For an extensive comparison of these paradigms including ours see the review of Liu et al. (2016). A second critical aim of the MIC was the assessment of distinct age effects. The model of neurocognitive aging by Wilson, Gallagher, Eichenbaum, and Tanila (2006) suggests that the aged memory system should show a bias toward pattern completion and a concurrent deficit in pattern separation. Consistent with this prediction, pattern separation has been shown to be less efficient in aging (Toner, Pirogovsky, Kirwan, & Gilbert, 2009), accompanied by CA3/DG hyperactivity and rigidity (Yassa et al., 2011a, 2011b), and it is more sensitive to cognitive decline than standard recognition memory (Holden, Toner, Pirogovsky, Kirwan, & Gilbert, 2013; Stark, Yassa, Lacy, & Stark, 2013). Complementary, our previous results suggested deficient but reinforced pattern completion (Vieweg et al., 2015). Thus, in a nutshell, the MIC can identify (a) pattern completion ability and (b) potential (age-related) response biases respectively favoring or restraining pattern completion.

In addition to behavioral performance, viewing behavior can inform about memory-related processing. For example, higher fixation numbers during encoding have been suggested to reflect increased accumulation of information, leading to better memory representations (Pertzov, Avidan, & Zohary, 2009) and hence increased memory performance (Loftus, 1972). This can be important when interpreting behavioral results in pattern separation/completion tasks, which have attributed age-related performance changes to a shift from pattern separation to pattern completion during memory retrieval (Yassa, Mattfeld, et al., 2011). However, Molitor et al. (2014) found fewer fixations on falsely recognized lure items during encoding, suggesting that a purely retrieval-based process is unlikely or at least insufficient to explain an observed behavioral bias.

To exhaust the possibilities of a task specifically tailored to pattern completion processes, we sought to refine and expand the MIC. First, we wanted to replicate the previous findings and rule out contributions of perceptual confounds. Even though the paradigm does not use incidental encoding and accommodates potential encoding differences through a learning criterion, we investigated whether specific viewing patterns could explain any age-related retrieval differences. This question was addressed with an eye-tracking study that characterized eye movements associated with the MIC and their relation to cognitive aging.

Second, for the MIC to be useful in longitudinal studies and for clinical research, it needs to be short, and parallel versions should be available. Therefore, in a second experiment we developed four shorter versions of the MIC and assessed performance stability over multiple testing occasions.

2 | EXPERIMENT 1

2.1 | Methods

The task administered here parallels the MIC task we have used in our previous work and which is described in more detail there (Vieweg et al., 2015). Concurrently, we used eye-tracking to monitor participants' viewing behavior during the task.

2.2 | Subjects

All participants were recruited from existing databases at the German Center for Neurodegenerative Diseases (DZNE) in Magdeburg. They underwent several neuropsychological tests and health assessments prior to the experiment: health questionnaire (Diersch, 2013), Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), multiple choice vocabulary test (MWT-B; Lehrl, Triebig, & Fischer, 1995), digit symbol substitution test (Wechsler, 2008), Rey–Osterrieth complex figure (ROCF) test as copy and 30 min delayed recall (DR; Rey, 1941; Corwin & Bylsma, 1993); their visual dominance was assessed, as well as subjective eyesight (using German grading system 1 "very good"–6 "insufficient") and objective eyesight (visual acuity determined on a pocket card test). Participants who scored less than 23 on the MoCA (N = 2; Luis, Keegan, & Mullan, 2009), who had chronic psychological or neurological problems (N = 1), who had insufficient eyesight

(N = 3), or whose eyes could not be tracked because of their glasses (N = 4; e.g., reflection, small frame, and dark frame), were excluded from the study. The remaining 26 young (21–35 years; 13 females; 6 wore glasses and 7 wore contact lenses) and 24 older adults (63–77 years; 12 females; 22 wore glasses) were healthy and had neuropsychological scores characteristic of their age group, including increased vocabulary scores and reduced DR performance (see Table 1). These findings are common in aging studies (Lövdén, Schellenbach, Grossman-Hutter, Krüger, & Lindenberger, 2005) and authenticate the cognitive typicality of our sample.

Informed written consent was obtained before the experiment, and the study received approval from the Ethics Committee of the University of Magdeburg. All participants received monetary compensation of 6.50 €/hr.

2.3 | Materials

We used 15 black and white line-drawn images of simple indoor scenes from Hollingworth and Henderson (1998; 5 to be learned, 5 to test learning, and 5 new). Stimulus completeness was manipulated by gradually masking all stimuli used in the test phase with a grid of white circles resulting in five different completeness levels (100%, 35%, 21%, 12%, and 5%; percentages reflect the amount of the image visible through the mask; see Figure 1(a)). Including the complete stimuli, this resulted in a total of 50 stimuli in the test phase (10 images × 5 completeness levels), each of which was repeated four times.

2.4 | Procedure

In a short study phase, participants had to learn a set of five different line-drawn scenes. Each item was presented randomly with a preceding semantic stimulus label. After learning, we verified that participants had learned the items by testing whether they could correctly identify them among new items until reaching a learning criterion (three consecutive correct identifications). In the main test phase of the experiment, participants had to consecutively judge the identity of a given stimulus out of a fixed set of response options (all of the learned stimuli's labels: "bar", "library", "dining room", "bedroom", "kitchen" and "none of these"; chance level: 1/6). More frequent applicability of the last option does not induce an overall response tendency to favor "none of these" responses (see Supporting Information for response distribution and discussion). A presented stimulus could either be a learned or a new stimulus, and was either presented in full (100%) or in a gradually masked version (35%, 21%, 12%, and 5%). Each stimulus was shown four times and for 2 s in pseudorandom order, so that no stimulus could be followed by a stimulus of the same identity or completeness level. Responses were self-paced, and participants were asked to respond accurately rather than quickly. Responses were scored as correct only when participants had picked the correct name for a learned stimulus or "none of these" for a new stimulus.

Additionally, eye movements were recorded during image presentation in all parts of the experiment. At the beginning of each phase of the experiment, gaze position was calibrated with a ninefold grid of fixation points. Participants were asked to refrain from looking down at the keyboard when responding, because head movements could cause slight shifts of the eye-tracker. However, as participants had to perform several button presses with at least five response options per trial, this could not be prevented entirely. Therefore, every trial started with a drift correction prior to image presentation, in which participants had to fixate a small white circle in the middle of the screen. If drift correction exceeded a visual angle of 5°, tracking of gaze position was adjusted by recalibration. During the test phase, gaze position was recalibrated every 70 trials by default. Gaze position was always recorded from the dominant eye (32 right and 18 left), except for two participants, for whom the recorded eye was switched after half of the experiment because of recalibration failure. Eye-movement data of two young participants were excluded for the first half of the study phase because of a calibration error.

2.5 | Bias measure

In our previous article (Vieweg et al., 2015), we have established a bias score that assesses whether participants lean toward either pattern completion or pattern separation. A more detailed rationale and description is available there, but we will shortly outline its definition in the following paragraph.

If pattern completion and separation are at balance, one should be just as good at recognizing learned items as at identifying novel items. The symmetry between these processes can shift, for example, because of age-related changes in hippocampal dynamics (for a detailed model, see Wilson et al., 2006), the behavioral consequence of which we wanted to illustrate. Given our focus on pattern completion, we calculated two scores: First, performance for learned items served as an index for individual *pattern completion ability*, with larger scores indicating a better ability to recognize familiar stimuli. Second, to quantify the *pattern completion bias*, we subtracted performance for new items from the performance for learned items. The deviation

TABLE 1 Experiment 1—Health questionnaire and neuropsychological data

Age group	School (years)	Higher education (years)	MoCA score	MWT-B		Digit symbol substitution		ROCF test		Eyesight	
				Raw	%	Raw	Score	Copy	30 min DR	Subjective	Objective
Young	12.83 (0.89)	4.42 (2.81)	28.36 (0.93)	30.72 (2.66)	75.98 (16.47)	88.58 (17.51)	12.0 (3.16)	35.16 (1.44)	25.06 (5.82)	1.71 (0.75)	0.71 (0.1)
Old	10.63 (1.34)	5.09 (3.22)	26.96 (1.99)	32.96 (1.51)	91.84 (8.45)	56.88 (10.84)	11.13 (1.83)	34.35 (2.37)	18.02 (4.3)	2.04 (0.51)	0.63 (0.12)

Note. Mean values (standard deviation). subjective eyesight: rated using German grading system 1 (very good)–6 (insufficient); objective eyesight: visual acuity determined on a pocket card test. Abbreviations: DR, delayed recall; MoCA, Montreal Cognitive Assessment; MWT-B, multiple choice vocabulary test; ROCF, Rey–Osterrieth complex figure.

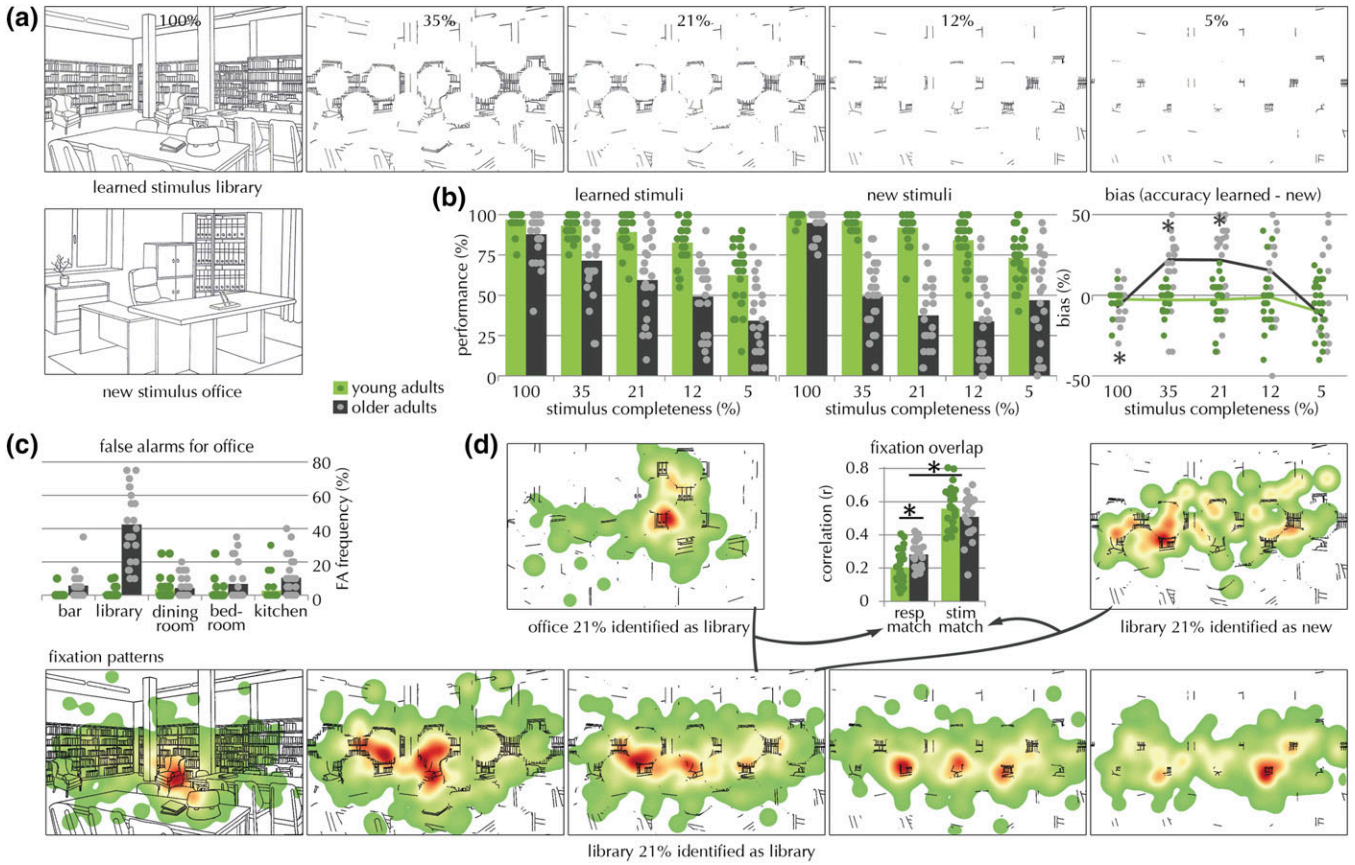

FIGURE 1 Experiment 1. (a) Learned stimulus library in all possible levels of completeness (percentages indicate portions of the image that are still visible through the mask) and new stimulus office. (b)Left: performance for both age groups (green: young adults and gray: older adults), separately for learned and new stimuli for the five different levels of stimulus completeness (mean); right: bias measure (see methods for a detailed explanation)—difference in accuracy scores for learned minus new stimuli calculated separately for each participant (mean); positive values indicate a bias toward pattern completion, significant differences from 0 are indicated with * separately for each age group as indicated by color. (c) Exemplary false alarm distribution for new stimulus office showing that older adults chose one particular false response option (library) most often rather than guessed more overall, which would have led to similar frequencies for all five response options. (d) Exemplary fixation heatmaps, warmer colors indicate higher fixation numbers; lower panel: fixation heatmaps for correctly identified stimuli (exemplary: library identified as library) for each masking level (100–5%); top left: fixation heatmap for a response-matching erroneous response (exemplary: office 21% identified as library), top right: fixation heatmap for a stimulus-matching erroneous response (exemplary: library 21% identified as new); top middle: fixation overlap—fixation heatmaps for correctly identified stimuli were correlated with heatmaps for stimulus-matching erroneous responses (stim match), or with response-matching erroneous responses (resp match). Pearson's correlation coefficients were Fisher's *z*-transformed and averaged across stimuli and completeness levels, separately for both age groups (correlation values in the figure were back-transformed for better comprehensibility), significant differences are indicated with *: stim-match correlations were overall higher than resp-match correlations suggesting that the viewing patterns were driven by the stimulus rather than by the response choice, but older adults' fixation patterns for resp-match stimuli correlated more than those of young adults (and these fixation overlap scores correlated with the response bias in older adults; see Section 3 for details) suggesting that the more likely a participant is to falsely recognize the office as the library, the more the fixation patterns for library and office overlap [Color figure can be viewed at wileyonlinelibrary.com]

from 0 reveals if participants were better at identifying learned items (positive score, indicative of a bias toward pattern completion) or new items (negative score, indicative of a bias toward pattern separation). The identification of new items (especially when they are incomplete) likely requires pattern separation as the scarcity of the information in the stimuli necessitates comparison and orthogonalization to learned stimuli, but see our previous article for a more extensive explanation (section 2.4 in Vieweg et al., 2015).

2.6 | Eye-tracking acquisition and analyses

Eye movements were recorded with a head-mounted EyeLink II tracker (SR Research, ON, Canada) at a sampling rate of 500 Hz. Participants were seated 50 cm in front of a 15 in. computer screen (1,024 × 768 pixel resolution, 60 Hz refresh rate). The experiment was programmed in Matlab 2013a with the Psychtoolbox Add-on to integrate the eye-tracker. Eye movements were recorded for each participant, and saccade, blink, and fixation data were calculated as follows. Blinks were defined as missing pupil data over three consecutive samples. Fixations were defined as static gaze positions, that is, no movements greater than 0.1° visual angle, faster than 30°/s velocity, and 8,000°/s^2 acceleration. Saccades consisted of all other recordings. Eye-tracking analyses focused on number and duration of fixations during the 2 s image presentation at study and test, and their proportion in predefined interest areas (IAs). IAs were equivalent to the inverse masking grid used to manipulate stimulus completeness,

that is, they consisted of all the areas in the image that were still visible through the mask, including an extended margin of 25 pixels to include areas that were in the foveal focus (~1.5° visual angle).

To investigate the spatial distribution of fixations, fixation heatmaps were extracted per trial for each participant, that is, a matrix containing fixation numbers at each image pixel coordinate. Each fixation's foveal area was also included into the fixation matrix; it was defined as ~1.5° visual angle around the fixation (25 pixel radius). Heatmaps were then averaged across repetitions and smoothed with a Gaussian filter (50 × 50 pixels), resulting in one fixation heatmap per condition (accounting for stimulus identity, completeness and response made, see Figure 1(d)). Two-dimensional Pearson's correlation coefficients were then calculated between condition-specific heatmaps.

2.7 | Exploration of group differences using Bayesian analyses

Complementary to standard frequentist statistics (*t*-tests and analysis of variance [ANOVA]), we performed Bayesian analyses to estimate support for the null hypotheses, because the absence of a significant difference between two measures in classical frequentist statistics cannot directly be taken to indicate that the two measures are in fact not different. Consequently, we calculated Bayes factors in JASP Version 0.8.4.0 (jasp-stats.org) using Bayesian *t*-tests and ANOVAs with default settings (*t*-test: Cauchy prior width 0.707; ANOVA: prior fixed effects 0.5, random effects 1, covariates 0.354) providing an indication of the strength of evidence in favor of the null hypothesis, that is, if there were no differences between age groups. The Bayes factor comparing the null hypothesis to the alternative hypothesis (BF_{01}) means that the data are *BF* times more likely under the null than the alternative hypothesis. BF_{01} values much greater than 1 support the conclusion that there is strong evidence in favor of the null hypothesis. We report BF_{01} only (and not BF_{10}) because our main focus was on the likelihood of the groups being similar. Note, though, that the null hypothesis is inherently more difficult to prove than the alternative, that is, with increasing observations the *BF* converges toward the true hypothesis (null or alternative), but to reach moderate (*BF* > 3) or strong (*BF* > 10) evidence, considerably more data are required than for the alternative (see minimal *BF* in Felix Schönbrodt's blog entry: http://www.nicebread.de/what-does-a-bayes-factor-feel-like/).

3 | RESULTS

3.1 | Behavioral results

3.1.1 | Accuracy

The following results were obtained by a three-way mixed ANOVA with factors age (young/old), stimulus completeness (100%, 35%, 21%, 12%, and 5%), and stimulus type (learned/new). Young participants performed better than older participants (main effect of age: $F_{[(1, 48)}$ = 142.01 p < .001). Reduced stimulus completeness resulted in less accurate performance (main effect of stimulus completeness: $F_{(4,192)}$ = 153.102, p < .001), and this decrease was more pronounced in older adults (age × stimulus completeness: $F_{(4,192)}$ = 33.582, p < .001). Performance per stimulus type did not differ overall (main effect of stimulus type: $F_{(1,48)}$ = 0.568, p = .455), but interactions revealed that older participants were less accurate for new stimuli compared with learned stimuli (age × stimulus type: $F_{(1,48)}$ = 5.408, p = .024; age × stimulus type × stimulus completeness: $F_{(4,192)}$ = 14.582, p < .001). Post hoc *t*-tests showed age group differences in performance across all levels of stimulus completeness for both learned and new stimuli (after Holm–Bonferroni multiple comparisons correction; all p < .01; level 100%: $t_{learned(48)}$ = 2.887, $t_{new(48)}$ = 2.936, level 35%: $t_{learned(48)}$ = 4.667, $t_{new(48)}$ = 11.045, level 21%: $t_{learned(48)}$ = 5.676, $t_{new(48)}$ = 12.418, level 12%: $t_{learned(48)}$ = 6.508, $t_{new(48)}$ = 9.196, level 5%: $t_{learned(48)}$ = 4.774, $t_{new(48)}$ = 4.024). In summary, both age groups' recognition ability decreased with reduced stimulus completeness, and older adults' recognition ability was impaired in comparison to young adults, especially for new stimuli (see Figure 1(b)).

3.1.2 | Response bias

We investigated potential response biases by subtracting individual accuracy scores for new stimuli from those for learned stimuli (see also Vieweg et al., 2015). Positive scores are indicative of better performance for to-be-retrieved (learned) stimuli suggesting a bias toward pattern completion, where as negative scores indicate better performance for to-be-encoded (new) stimuli in turn suggesting a bias toward pattern separation.

A mixed ANOVA (age × stimulus completeness) revealed that stimulus completeness influenced the response bias (main effect of stimulus completeness: $F_{(4,192)}$ = 14.582, p < .001), and older adults showed a more positive response bias than young adults (see Figure 1 (b), right; main effect of age: $F_{(1,48)}$ = 5.408, p = .024). A two-way interaction indicated that stimulus completeness differentially affected the response bias dependent on the age group (age × stimulus completeness: $F_{(4,192)}$ = 8.053, p < .001). Follow-up *t*-tests demonstrated between-group differences for the middle three completeness levels, but after Holm–Bonferroni multiple comparisons correction the 12% level did not reach significance (level 35%: $t_{(48)}$ = −4.68, p < .001; level 21%: $t_{(48)}$ = −4.042, p < .001; level 12%: $t_{(48)}$ = −2.107, p = .04).

We examined at which levels of stimulus completeness the scores established a response bias. To that end, per group, 5 one-sample *t*-tests of the average response bias scores against 0 revealed that older adults showed a positive response bias for completeness levels 35% and 21%, indicative of a pattern completion bias, and a negative response bias for level 100% after Holm–Bonferroni multiple comparisons corrections (see Figure 1(b); level 100%: $t_{(23)}$ = −2.416, p = .024; level 35%: $t_{(23)}$ = 4.149, p < .001; level 21%: $t_{(23)}$ = 3.819, p = .001; level 12%: $t_{(23)}$ = 2.081, p = .049; level 5%: $t_{(23)}$ = −1.392, p = .177). Young participants showed no response bias at all (all p > .05).

We did not purposefully select learned and new stimuli for matching similarity; in fact, we explicitly refrained from manipulating stimulus similarity to not increase pattern separation demands. However, if participants completed toward the stimulus perceived as most similar, one false response option should have been chosen over all

other false options. Alternatively, if participants were merely guessing, all false response options should occur equally often. Therefore, we investigated older adults' distribution of false alarm options for new stimuli by calculating the group-average frequencies of all false choice options. Frequencies were then sorted from most (FA 1) to least (FA 5) chosen option for each new stimulus, and averaged across all new stimuli afterwards. A χ^2-test of goodness-of-fit on the 5 false alarm options ($\chi^2_{(4)}$ = 396.121, p < .001) revealed that participants indeed chose one particular false response option most often and were not randomly guessing (see Table 2). We inspected whether there was one dominant option for each stimulus, by contrasting the most frequent false alarm against the average of the other false alarms per stimulus (see Figure 1(c) for one example; stimulus "office": $\chi^2_{(1)}$ = 126.863, p < .001; stimulus "class room": $\chi^2_{(1)}$ = 19.862, p < .001; stimulus "restaurant": $\chi^2_{(1)}$ = 16.03, p < .001; stimulus "locker room": $\chi^2_{(1)}$ = 11.449, p = .001; stimulus "living room": $\chi^2_{(1)}$ = 2.462, p = .117). This indicates that older participants mostly completed toward the stimulus perceived as most similar.

3.1.3 | Summary of the performance results

The present results replicated the findings from our previous study (Vieweg et al., 2015); both age groups' recognition memory declined with decreasing stimulus completeness, older adults showed a stronger decline and they chose familiar responses over new ones, suggesting a bias toward pattern completion. Note that reaction times and confidence ratings followed the same profile as the original study (see Supporting Information).

3.2 | Eye-tracking results

3.2.1 | Fixations during encoding

To investigate whether behavioral performance in the MIC could be driven by encoding effects, we examined fixations during the study phase. Fixation durations did not differ between age groups ($t_{(48)}$ = 0.715, p = .478, BF_{01} = 2.867), but younger adults had slightly higher fixation numbers than older adults ($t_{(48)}$ = 2.053, p = .048, BF_{01} = 0.650; see left panel in Figure 2). In order to explore whether this difference was responsible for the age-related response bias during retrieval, we correlated the number of fixations during encoding with a cumulative partial response bias score per participant (mean of all partial response biases; as in Baker et al. (2016)). In neither of the age groups did these measures correlate (Pearson correlation; young adults: r = –0.282, p = .163; older adults: r = –0.090, p = .674) suggesting that encoding differences were not predictive of response biases. Similarly, the older adults' reduced recognition of learned items was not driven by differential encoding either (Pearson correlation with cumulative performance for partial learned items; young adults: r = 0.073, p = .721; older adults: r = 0.060, p = .758).

3.2.2 | Fixations on the whole image during retrieval

First, we inspected the overall number of fixations in a similar way as for behavioral accuracy. We subjected the fixation numbers to a three-way mixed ANOVA with factors age (young and old), stimulus completeness (100%, 35%, 21%, 12%, and 5%) and stimulus type (learned and new). Interestingly, young and older adults did not differ in how much they fixated on the images (see middle panel in Figure 2; main effect of age: $F_{(1,48)}$ = 0.604, p = .441, BF_{01} = 2.184), nor was there a difference between learned and new stimuli (main effect of stimulus type: $F_{(1,48)}$ = 0.480, p = .492, BF_{01} = 8.150). However, participants fixated slightly more when less of the image was visible (main effect of stimulus completeness: $F_{(4,192)}$ = 11.083, p < .001, BF_{01} = 0.086).

Second, there was an inverse relationship between the number of fixations and their average duration, that is, the longer participants spent fixating on one location, the lower the number of fixations was and vice versa. Given that participants only had 2 s to look at each image, this reflects a trade-off between how frequently to shift fixations and how long to hold each of them. Therefore, we analyzed the sum of durations for all fixations rather than average fixation duration. While there were no main effects for age or stimulus type (main effect of age: $F_{(1,48)}$ = 3.189, p = .08, BF_{01} = 0.823; main effect of stimulus type: $F_{(1,48)}$ = 0.442, p = .510, BF_{01} = 8.855), stimulus completeness affected fixation durations (main effect of stimulus completeness: $F_{(4,192)}$ = 24.035, p < .001, BF_{01} = 1.170e − 20), and interacted with age (age × stimulus completeness: $F_{(4,192)}$ = 8.671, p < .001). More specifically, older adults fixated only shortly on the full stimuli but spend more time fixating the masked stimuli with durations similar to those of young adults whose fixation durations did not differ across masking levels.

3.2.3 | Fixations in IAs during retrieval

To identify areas of the images at which participants were looking, we examined fixations in predefined IAs. As explained in the methods, the stimuli were masked at different levels, so that only parts of the image were visible. We tested whether the age groups showed differential fixation patterns at the relevant positions, that is, the parts of the image that still carried information, which might explain the observed age-related response bias. Thus, IAs comprised all the unmasked portions of the image with an added margin of 25 pixels (to include areas in the foveal focus).

Again, as revealed by a three-way mixed ANOVA, young and older adults did not differ overall in their number of fixations on IAs (see right panel in Figure 2; main effect of age: $F_{(1,48)}$ = 1.497, p = .227, BF_{01} = 2.092), nor were there differences between learned and new stimuli (main effect of stimulus type: $F_{(1,48)}$ = 0.288, p = .594, BF_{01} = 9.522). Both groups had less fixations on less complete images (main effect of stimulus completeness: $F_{(4,192)}$ = 66.497, p < .001, BF_{01} = 3.513e − 59). That was to be expected, because the IAs got smaller with increasing masking levels. The same was true for fixation

TABLE 2 Experiment 1—Response distribution for new stimuli

Age group	CR	FA1	FA2	FA3	FA4	FA5
Young	0.89 (0.07)	0.04 (0.04)	0.02 (0.02)	0.03 (0.03)	0.01 (0.02)	0.01 (0.01)
Old	0.43 (0.2)	0.26 (0.14)	0.13 (0.07)	0.09 (0.05)	0.06 (0.05)	0.05 (0.03)

Note. Mean values (standard deviation). Abbreviations: CR, correct rejection; FA, false alarm.

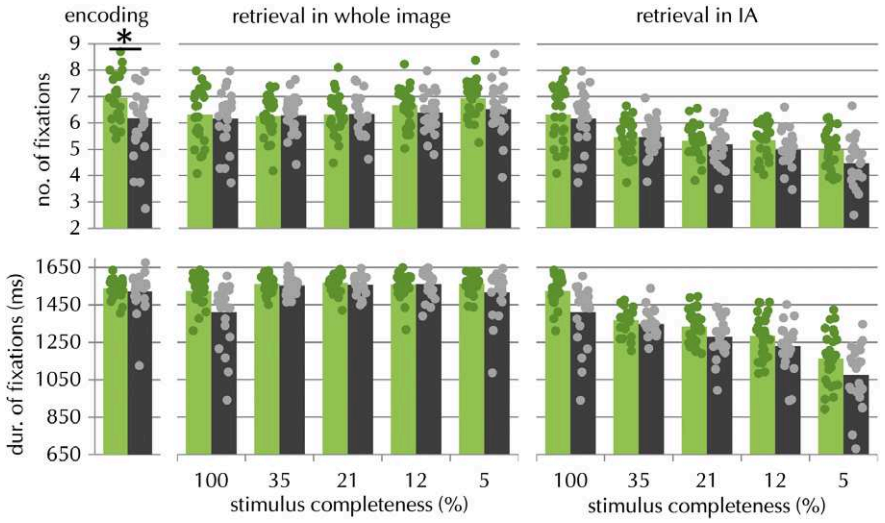

FIGURE 2 Experiment 1. Fixation numbers and durations (mean). Left, during encoding, significant effects are indicated with *; middle, during retrieval in the whole image; right, during retrieval within specific IAs (the parts of the image that still carry information). Data are presented for both age groups (green: young adults and gray: older adults) and all stimulus completeness levels, but collapsed over learned and new stimuli as there were no differences between them; sum of durations in ms [Color figure can be viewed at wileyonlinelibrary.com]

durations, that is, all participants fixated shorter on smaller IAs and spent more time fixating on the more complete stimuli (main effect of stimulus completeness: $F_{(4,192)} = 125.125$, $p < .001$, $BF_{01} = 7.017e-88$), and there were no differences between learned and new stimuli (main effect of stimulus type: $F_{(1,48)} = 2.223$, $p = .143$, $BF_{01} = 7.427$). However, here younger adults fixated slightly longer on the images than older adults (main effect of age: $F_{(1,48)} = 6.826$, $p = .012$, $BF_{01} = 0.343$).

3.2.4 | Spatial distribution of fixations

In addition to investigating the number and duration of fixations, we also tested whether different viewing patterns could drive recognition performance and the corresponding age differences. Specifically, we wanted to investigate whether recognition errors coincided with altered viewing behavior compared with eye-movement patterns during correctly identified trials. We assumed that each stimulus would elicit viewing patterns specific to that stimulus when recognized. However, we wanted to know if this pattern differed when stimuli where not recognized. Given that particularly the new stimuli were often erroneously recognized as one of the learned stimuli by older adults, we hypothesized that viewing patterns for these stimuli could be similar to the stimulus participants thought they were seeing. In contrast, viewing patterns could be more driven by basic visual features of each stimulus independent of participants' recognition. Therefore, we computed the spatial overlap of fixations between conditions. Specifically, we calculated two-dimensional correlation coefficients between fixation heatmaps (see methods eye-tracking analysis) of correctly identified learned stimuli (e.g., library identified as library) with falsely identified stimulus-matching stimuli (e.g., library identified as new) and with falsely identified response-matching stimuli (e.g., office identified as library), separately for all levels of stimulus completeness (for exemplary heatmaps and how the correlations were calculated, see Figure 1(d)). We calculated correlations of all possible connected stimulus–response pairs, for example, for library trials erroneous stimulus-matches could be library identified as new, bar, dining room, bedroom, or kitchen; erroneous response-matches could be bar, dining room, bedroom, kitchen, or all new stimuli that were identified as library. The resulting r values were Fisher's z-transformed and collapsed across stimulus identity, response and completeness, resulting in one response-matching and one stimulus-matching correlation value per participant. To identify differential viewing patterns pertaining to these conditions and assess potential age differences, the correlation values were subjected to a two-way mixed ANOVA with factors age group and match type. There were no differences between age groups (main effect of age: $F_{(1,48)} = 0.002$, $p = .965$, $BF_{01} = 4.166$), but correlations for stimulus-matching heatmaps were considerably higher than for response-matching heatmaps (main effect of match type: $F_{(1,48)} = 194.381$, $p < .001$, $BF_{01} = 5.074e-21$), suggesting that the viewing patterns were driven by the stimulus rather than by the response choice. This effect also interacted with age (age × match type: $F_{(1,48)} = 9.332$, $p = .004$). Post hoc t-tests revealed that older adults' fixation patterns for response-matching stimuli correlated more than those of young adults ($t_{(48)} = -2.882$, $p = .006$, $BF_{01} = 0.136$), but there were no group differences between stimulus-matching heatmap correlations ($t_{(48)} = 1.512$, $p = .137$, $BF_{01} = 1.397$). To test the heatmaps' relation to the suggested pattern completion bias, we correlated the heatmap match scores with the cumulative partial response bias (Pearson correlation). Both the stimulus-matching and the response-matching score positively correlated with the response bias in older adults (stimulus-matching: $r = 0.483$, $p = .017$; response-matching: $r = 0.418$, $p = .042$), but not in young adults (stimulus-matching: $r = 0.005$, $p = .981$; response-matching: $r = 0.346$, $p = .084$), indicating that the overlap of fixation heatmaps was related to the age-related pattern completion bias.

3.2.5 | Summary of the eye-tracking results

During encoding, older adults had slightly lower fixation numbers than young adults, however, this difference was not predictive of the observed age-related response bias. Both age groups had similar fixation patterns during retrieval albeit some differences in fixation

durations. Specifically, older adults had shorter viewing durations than young adults for complete stimuli and within IAs. Given the appearance of a masked stimulus, one may argue that the observed viewing pattern was a useful strategy for this task. That is, with decreased stimulus completeness there was less to see at each fixation, rendering it necessary to shift fixations more often to obtain a similar amount of information as for the unmasked stimuli. Importantly, there were no age-related viewing differences between learned and new stimuli which could account for the observed response bias. Analyses of fixation heatmaps revealed that viewing patterns were more stimulus-driven than predictive of response choices. That is, independent of recognition, fixation patterns for example, "library" trials were more similar to other "library" trials than fixation patterns for "office" trials that were falsely recognized as "library". Interestingly though, within the response-matching heatmap comparison, older adults had higher values than young adults, which also correlated with the response bias. These results suggest that the spatial overlap between fixations was related to the response bias in older adults, that is, the more fixation patterns overlap between correctly and incorrectly identified stimuli, the stronger the bias to choose a familiar response and to falsely recognize a new stimulus.

4 | EXPERIMENT 2

4.1 | Methods

To obtain longitudinal performance profiles and detect potential performance changes in clinical studies, it is necessary to test participants on multiple occasions. We therefore aimed at developing parallel versions of the MIC, comparable in difficulty. We used fewer stimuli to facilitate quick administration often necessary in clinical studies. In addition, a subset of participants was tested twice on different versions of the task to assess the reliability of the task and to investigate potential learning effects independent of the stimuli.

4.2 | Subjects

All participants were recruited from existing databases at the DZNE in Magdeburg. Sixty young and healthy adults (19–35 years; 43 females) participated in Experiment 2, and were randomly assigned to one of four groups pertaining to a specific version of the task (15 participants per task version). Thirty-six of the participants came in for a second time and were again distributed across versions (nine participants per task version). Informed written consent was obtained before the experiment, and the study received approval from the Ethics Committee of the University of Magdeburg. All participants received monetary compensation of 6.50 €/hr.

4.3 | Materials

Forty-eight black and white line-drawn images of simple indoor scenes were created using Autodesk 3DS Max. Four different sets of 12 stimuli each (4 to be learned, 4 to test learning, and 4 new) constituted different versions of the task (see exemplary stimuli for two sets Figure 3(a)). Again, the stimuli were gradually masked (stimulus completeness: 100%, 35%, 21%, 12%, and 5%). In total, there were 40 stimuli in the test phase of each version, and each stimulus was repeated four times.

4.4 | Procedure

We administered the MIC as described in Experiment 1 but in four different versions. Apart from containing a different set of stimuli, all task versions were identical. Fifteen participants each were randomly assigned to one of the four versions. In each version, participants had to learn four different scene exemplars, resulting in a chance level of 1/5 in the test phase. As in Experiment 1, there was no overall response tendency to favor "none of these" responses as a result of the response option distribution (see Supporting Information). Thirty-six of the participants performed two versions of the task with a minimum of 2 weeks in-between tests. All possible pairings of tasks were tested in a between-subjects design (e.g., versions 1–2, 1–3, 1–4, 2–1, 2–3, etc.).

4.5 | Exploration of task differences using Bayesian analyses

As in Experiment 1, we report Bayes factor BF_{01} in favor of the null hypothesis to estimate support for the similarity of the task versions.

5 | RESULTS

5.1 | Results at initial testing

In order to compare the four different versions, performance data of all initial tests (N = 60) were submitted to a three-way mixed ANOVA with between-subjects factor version (1–4), and within-subjects factors stimulus completeness (100%, 35%, 21%, 12%, and 5%) and stimulus type (learned and new). Performance across all four versions differed significantly (see Figure 3(b); main effect of version: $F_{(3,56)}$ = 3.935, p = .013). Reduced stimulus completeness resulted in less accurate performance (main effect of stimulus completeness: $F_{(4,224)}$ = 102.147, p < .001), and performance was better for new than for learned stimuli (main effect of stimulus type: $F_{(1,56)}$ = 35.826, p < .001). Additionally, there was an interaction between version and stimulus completeness but no other interactions were significant (version × stimulus completeness: $F_{(12,224)}$ = 2.616, p = .003; version × stimulus type × stimulus completeness: $F_{(12,224)}$ = .911, p = .537, version × stimulus type: $F_{(3,56)}$ = 1.949, p = .123). Importantly though, as analyzed by a separate ANOVA, response biases did not differ between versions (main effect of version: $F_{(3,56)}$ = 1.949, p = .132).

Descriptive data (see Figure 3(b)) showed that versions 3 and 4 yielded similar results, whereas version 1 seemed to be easier and version 2 seemed to be more difficult. As we wanted to obtain comparable versions, we continued analyses only with data from versions 3 and 4 (N = 30). Indeed, there was no overall difference between these two versions (see Figure 3(c) top panel; main effect of version: $F_{(1,28)}$ = 0.781, p = .384, BF_{01} = 3.586), nor were there any interactions with the other factors (version × stimulus type: $F_{(1,28)}$ = 0.098, p = .757; version × stimulus completeness: $F_{(4,112)}$ = 1.776, p = .139; version × stimulus type × stimulus completeness: $F_{(4,112)}$ = 0.569, p = .685). Other than that, participants were less accurate with decreasing

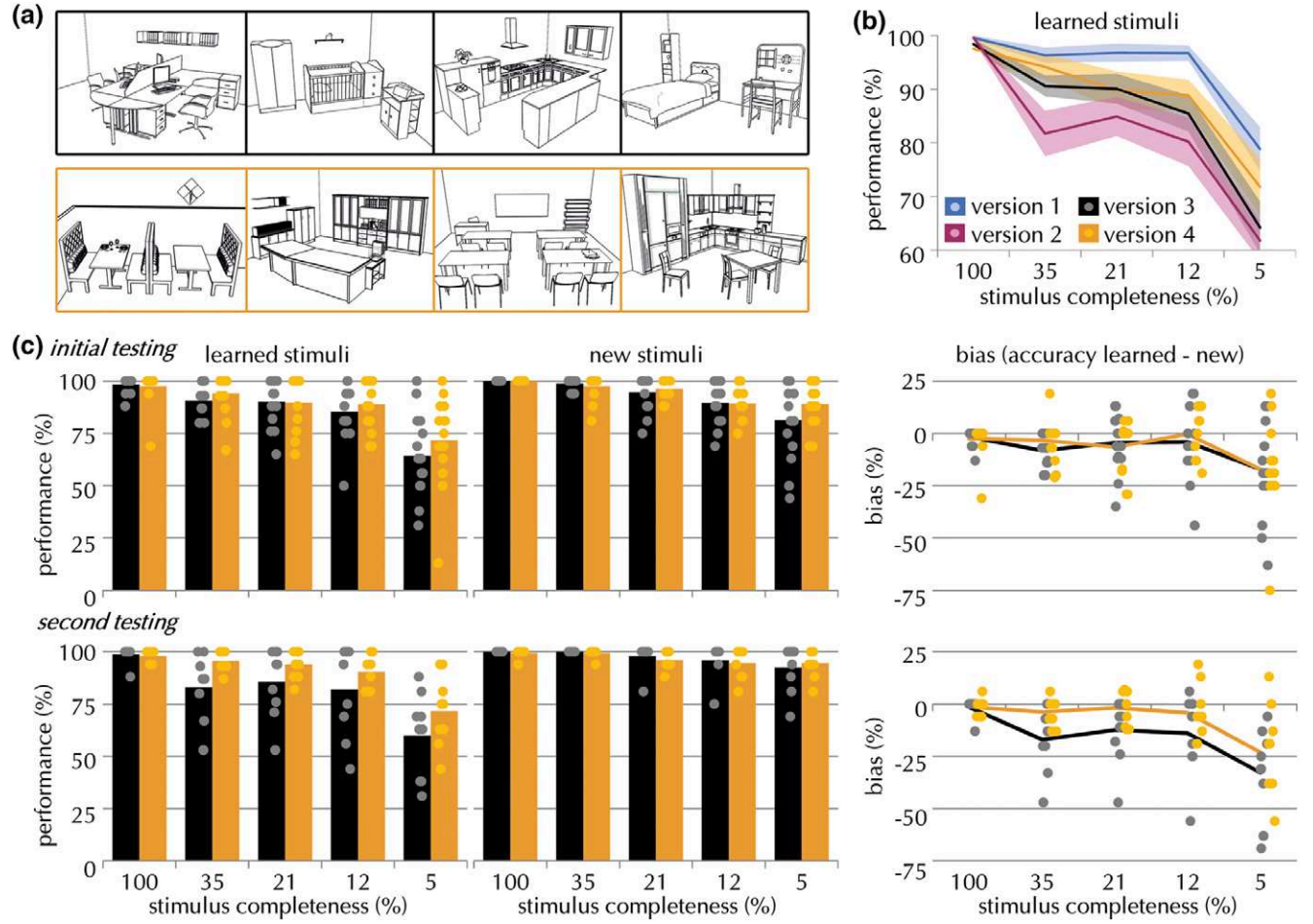

FIGURE 3 Experiment 2. (a) Exemplary learned stimuli for versions 3 (black) and 4 (orange). (b) Performance for learned stimuli separately for all tested versions (blue: version 1, magenta: version 2, black: version 3, and orange: version 4) and the five different levels of stimulus completeness (mean ± standard error). (c) Performance and bias measures at initial (top) and second testing (bottom) for comparable versions 3 and 4; left: performance separately for learned and new stimuli for the five different levels of stimulus completeness (mean); right, bias measure (see Section 2.5 for a detailed explanation)—difference in accuracy scores for learned minus new stimuli calculated separately for each participant (mean); positive values indicate a bias toward pattern completion [Color figure can be viewed at wileyonlinelibrary.com]

stimulus completeness (main effect of stimulus completeness: $F_{(4,112)} = 54.995$, $p < .001$, $BF_{01} = 4.351e-25$) and performed better for new than for learned stimuli (main effect of stimulus type: $F_{(1,28)} = 12.918$, $p = .001$, $BF_{01} = 7.407e-4$). Furthermore, the response bias also did not differ between versions (main effect of version: $F_{(1,28)} = 0.098$, $p = .757$, $BF_{01} = 3.242$; version × stimulus completeness: $F_{(4,112)} = 0.569$, $p = .685$), but it slightly decreased with decreasing stimulus completeness (main effect of stimulus completeness: $F_{(4,112)} = 9.646$, $p < .001$, $BF_{01} = 4.013e-5$). This was confirmed by one-sample t-tests against 0, revealing that there was a negative response bias at 5% stimulus completeness, but at no other completeness level (after Holm–Bonferroni multiple comparisons correction, level 100%: $t_{(14)} = -1.718$, $p = .108$; level 35%: $t_{(14)} = -1.624$, $p = .127$; level 21%: $t_{(14)} = -1.695$, $p = .112$; level 12%: $t_{(14)} = -0.569$, $p = .578$; level 5%: $t_{(14)} = -3.439$, $p = .004$).

5.2 | Results at second testing

Performance data of all participants who came in for a second test of versions 3 and 4 ($N = 18$) were submitted to a mixed ANOVA with the factors task version (3 and 4), stimulus completeness (100%, 35%, 21%, 12%, and 5%) and stimulus type (learned and new). Most importantly, the versions did not differ from each other (see Figure 3(c) bottom panel; main effect of version: $F_{(1,16)} = 1.634$, $p = .219$, $BF_{01} = 2.306$), nor were there any interactions (version × stimulus type: $F_{(1,16)} = 3.001$, $p = .102$; version × stimulus completeness: $F_{(4,64)} = 1.779$, $p = .144$; version × stimulus type × stimulus completeness: $F_{(4,64)} = 0.961$, $p = .435$). Reduced stimulus completeness resulted in less accurate performance (main effect of stimulus completeness: $F_{(4,64)} = 42.952$, $p < .001$, $BF_{01} = 4.742e-8$), and performance was better for new than for learned stimuli (main effect of stimulus type: $F_{(1,16)} = 19.776$, $p < .001$, $BF_{01} = 3.921e-7$). Additionally, response biases did not differ between versions (main effect of version: $F_{(1,16)} = 3.001$, $p = .102$, $BF_{01} = 1.083$; version × stimulus completeness: $F_{(4,64)} = 0.961$, $p = .435$), but they were affected by stimulus completeness (main effect of stimulus completeness: $F_{(4,64)} = 14.793$, $p < .001$, $BF_{01} = 6.334e-7$). Arguably though, BF_{01} for the factor version was around 1, indicating that no strong conclusion can be drawn about the version-specific difference of the response biases during second testing; descriptively, version 3 may show a slight negative bias in relation to version 4 (i.e., a bias toward

pattern separation; see Figure 3(c)), nevertheless, this is not backed up statistically.

5.3 | Comparison between time points

To be able to measure change over time, it is important that the parallel versions of the MIC are not sensitive to general practice or learning effects. Because of the between-subjects design across all four versions, we only had six participants who completed versions 3 or 4 at first and second testing. However, given that general learning effects should be independent of the particular version applied at the first testing session, we realized the time point analysis in a between-subjects design. Specifically, we compared performance between participants who were tested with versions 3 or 4 in the initial session (N = 15) with a different set of participants who were tested on versions 3 or 4 in the second session (N = 15). The six participants who had been tested on both versions were equally distributed over these two groups (three in each group; total per group N = 18). To keep group size equal, only participants who came in twice were included here.

A mixed ANOVA with the between-subjects factor time (initial or second test), and the within-subjects factors stimulus type and completeness revealed no overall learning effects between time points (see Table 3; main effect of time: $F_{(1,28)}$ = 1.408, p = .245, BF_{01} = 2.970), nor were there any interactions (time × stimulus type: $F_{(1,28)}$ = 0.389, p = .583; time × stimulus completeness: $F_{(4,112)}$ = 2.068, p = .090; time × stimulus type × stimulus completeness: $F_{(4,112)}$ = 0.637, p = .637). Effects of stimulus type and completeness followed the same profile as before (main effect of stimulus completeness: $F_{(4,112)}$ = 75.683, p < .001, BF_{01} = 2.748e − 23; main effect of stimulus type: $F_{(1,28)}$ = 21.807, p < .001, BF_{01} = 1.279e − 6). Importantly, response biases also did not change over time (main effect of time: $F_{(1,28)}$ = 0.389, p = .538, BF_{01} = 3.090; time × stimulus completeness: $F_{(4,112)}$ = 0.637, p = .637), but were again affected by stimulus completeness (main effect of stimulus completeness: $F_{(4,112)}$ = 17.317, p < .001, BF_{01} = 2.214e − 9).

5.4 | Summary of the results

Out of the four developed stimulus sets, versions 3 and 4 yielded comparable results which followed the same performance profiles as the original MIC, independent of whether they were tested for the first or a second time. We did not observe any learning effects when participants were tested twice on the task. Note that there were no differences in reaction times between versions or time points either (see Supporting Information).

6 | DISCUSSION

In this study, we could replicate the behavioral findings of our earlier article (Vieweg et al., 2015). Specifically, recognition memory declined with reduced stimulus completeness, more so in aging, and older adults chose familiar responses over new ones, suggesting a bias toward pattern completion. Associated eye-tracking data of older adults showed fewer fixations during encoding and slightly shorter IA fixation durations during retrieval but with no distinction between learned and new stimuli. While this particular difference may have contributed to the overall performance reduction of older adults, it could not account for the more specific impairment in the recognition of new items, that is, although the age groups showed small encoding differences, the age-related pattern completion bias could not be attributed to this difference. However, when a new stimulus was falsely identified as a learned stimulus, the corresponding fixation patterns overlapped more in older adults than in young adults, and only in older adults did the resulting score correlate with the performance bias suggesting a relation between the two. Nevertheless, spatial viewing patterns for both age groups were more driven by stimulus identity and not so much by response choice. Thus, the observed differential recognition memory effects cannot solely be explained by the corresponding eye-movements, lending support for our hypothesis that differences in pattern completion are accountable for the observed effects.

In a second experiment, we developed four distinct and shorter versions of the MIC to provide a means for repeated testing. Given the complexity of scene images, it is very likely that they differ in their recognizability, which is further influenced by the set within which they are situated. For example, the recognizability of the scene "kitchen" does not only depend on the image itself, but also on the visual similarity of other images from the same set of stimuli. Thus, it is far from surprising that different task versions differ in difficulty. Nevertheless, of the four tested versions, two tasks proved to be of similar difficulty, and performance for these two tasks was comparable across two time points, suggesting that there was no general learning effect of the task that could distort performance profiles. Consequently, we can now provide two task versions of comparable difficulty for repeated testing.

The previous literature on memory-related viewing behavior has linked increased fixation numbers during encoding to better memory retrieval (Loftus, 1972). With respect to that, older adults' fewer fixation numbers during encoding may be similarly related to their impaired retrieval—an issue that was also raised by Molitor et al. (2014) who found that false alarm trials were associated with worse encoding fixation patterns than correctly identified stimuli. However, here, correlations of eye-tracking data during encoding with pattern completion ability or bias were not observed. Consequently, performance differences could not be predicted from differential eye movements during encoding, suggesting the involvement of another process, which we suggest to be an age-related pattern completion bias.

Further on, fixation durations at informative positions (i.e., IAs) during retrieval were slightly longer in the younger age group, potentially accounting for their overall recognition advantage. This idea receives support from previous observations that longer fixations during retrieval would code for a prior exposure indicating recognition (Hannula, Baym, Warren, & Cohen, 2012; Molitor et al., 2014). However, reports in the literature also show that saccade velocity and reaction times (Moschner & Baloh, 1994) as well as saccade frequencies and amplitude (Dowiasch, Marx, Einhäuser, & Bremmer, 2015) decrease with age, in line with a general slowing of processing speed (O'Shea, Cohen, Porges, Nissim, & Woods, 2016). Based on these

TABLE 3 Experiment 2—Performance results for multiple timepoint testing per stimulus type and completeness levels

	Accuracy learned stimuli					Accuracy new stimuli					Bias (accuracy learned—new)				
Timepoint	**100%**	**35%**	**21%**	**12%**	**5%**	**100%**	**35%**	**21%**	**12%**	**5%**	**100%**	**35%**	**21%**	**12%**	**5%**
Initial test	0.96 (0.08)	0.93 (0.09)	0.89 (0.13)	0.85 (0.14)	0.61 (0.2)	1.0 (0.0)	0.97 (0.06)	0.94 (0.07)	0.88 (0.09)	0.83 (0.15)	−0.04 (0.08)	−0.04 (0.09)	−0.06 (0.12)	−0.03 (0.16)	−0.22 (0.24)
Second test	0.98 (0.04)	0.9 (0.12)	0.91 (0.09)	0.87 (0.13)	0.66 (0.17)	1.0 (0.02)	1.0 (0.02)	0.96 (0.05)	0.94 (0.08)	0.92 (0.09)	−0.02 (0.04)	−0.09 (0.13)	−0.05 (0.09)	−0.07 (0.13)	−0.26 (0.19)

Note. Mean values (standard deviation) combined for task versions 3 and 4.

findings, shorter fixation durations in older adults could be explained by their increased saccade durations, that is, eye-movements between fixations. Thus, shorter fixation durations and lower fixation numbers might well account for recognition memory deficits, however, they do not provide a reliable index of the specificity of the underlying process, be it processing speed or memory decline. Nevertheless, to prevent potential encoding differences related to processing speed in the future, we have adjusted our paradigm in such a way that during the learning phase participants can decide in each trial how long the stimulus is presented.

In further support of recognition rather than perceptual differences, the main age-related disparity in this task was observed in the difference in performance for learned and new stimuli, which was absent in the eye-tracking data. Hence, while older adults made more errors for new stimuli than for learned ones, their eye-movements were not differentiable between the two stimulus types. Nevertheless, the age group interaction for spatial overlap of fixations is noteworthy; although fixation patterns were generally more stimulus than response-driven, older adults' fixations for new stimuli—which were identified as old—shared more overlap with the corresponding old stimuli than was the case for young adults. This observation partly supports what we were assuming from the false alarm distributions, namely that for incorrect recognition, participants chose the stimulus they perceived as the most similar. However, given the distinctly fewer errors in young adults, their heatmap correlations consist of considerably fewer trials, which inevitably affect their overall correlation score, and consequently puts some limits to the interpretability of these scores. In summary, the specificity of the older adults' errors, that is, more often picking a specific familiar response instead of the "none of these" option in identifying new stimuli, and the absence of a clear corresponding viewing pattern, lends further credibility to the interpretation that age differences observed with this task are a result of a pattern completion bias.

It is also worth noting that the behavioral findings could almost completely be replicated from their first implementation (Vieweg et al., 2015). The only deviation was a nonsignificant group difference in the bias score for the 12% completeness level, but the overall shape of the curve was the same, and so were all other results concerning accuracies, false alarm distributions, reaction times, and confidence ratings (see Supporting Information for the latter two). This provides compelling evidence that the task in its current form can be reliably applied and the obtained results are not coincidental. Thus, the MIC can be used to pick up fine-grained age-related recognition memory differences in the scheme of pattern completion, which are in line with models of cognitive aging (Wilson et al., 2006). The MIC provides two important measures: pattern completion ability—quantified by recognition performance for learned stimuli and a pattern completion bias—quantified by the interplay between learned and new images.

Furthermore, we have tested and can now provide additional, parallel versions of the MIC whose performance profiles match the original task, but which are fast and easy to administer. The MIC, including all parallel versions, is publicly available on the open science framework (https://osf.io/juvwy) in order to advance the research field and to enable other scientists to employ a pattern-completion-specific task in their studies. We believe that the MIC could prove particularly useful for longitudinal studies (i.e., on lifespan development and dementia), or as a clinical tool in patient and intervention studies.

CONFLICT OF INTEREST

The authors declared that they have no conflict of interest.

ACKNOWLEDGMENTS

Parts of this article have been used in different form in P.V.'s doctoral thesis (Vieweg, 2017). P.V. and T.W. are supported by European Research Council Starting Investigator Grant AGESPACE (335090). M.R. is supported by European Social Fonds (ESF) (Sachsen-Anhalt Wissenschaft Spitzenforschung/Synergien: AGETIME). The authors would like to thank Henrike Raith, Franziska Schulze, Anica Luther, Patrick Hauff, and Sophie Schnau for their help with data collection and neuropsychological testing and Selim Candan for help with stimulus design.

SOFTWARE ACCESSIBILITY

The Memory Image Completion task (MIC) is available on the open science framework: https://osf.io/juvwy.

ORCID

Paula Vieweg https://orcid.org/0000-0003-2789-6970
Martin Riemer https://orcid.org/0000-0001-6043-4204
David Berron https://orcid.org/0000-0003-1558-1883
Thomas Wolbers https://orcid.org/0000-0003-2943-8949

REFERENCES

Baker, S., Vieweg, P., Gao, F., Gilboa, A., Wolbers, T., Black, S. E., & Rosenbaum, R. S. (2016). The human dentate gyrus plays a necessary role in discriminating new memories. *Current Biology*, *26*, 2629–2634.

Bakker, A., Kirwan, C. B., Miller, M. I., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*, 1640–1642.

Corwin, J., & Bylsma, F. W. (1993). Psychological examination of traumatic encephalopathy. *Clinical Neuropsychology*, *7*, 3–21.
Diersch, N. (2013). Action prediction in the aging mind. PhD Thesis, *Max Planck Institute for Human Cognitive and Brain Sciences*, Leipzig.
Dowiasch, S., Marx, S., Einhäuser, W., & Bremmer, F. (2015). Effects of aging on eye movements in the real world. *Frontiers in Human Neuroscience*, *9*, 46.
Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA1. *Neuron*, *44*, 581–584.
Hannula, D. E., Baym, C. L., Warren, D. E., & Cohen, N. J. (2012). The eyes know: Eye movements as a veridical index of memory. *Psychological Science*, *23*, 278–287.
Hasselmo, M. E., & McClelland, J. L. (1999). Neural models of memory. *Current Opinion in Neurobiology*, *9*, 184–188.
Holden, H. M., Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2013). Visual object pattern separation varies in older adults. *Learning & Memory*, *20*, 358–362.
Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398–415.
Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, *24*, 988–992.
Hunsaker, M. R., & Kesner, R. P. (2013). The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neuroscience & Biobehavioral Reviews*, *37*, 36–58.
Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory*, *14*, 625–633.
Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory*, *18*, 15–18.
Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, *91*, 335–345.
Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, *315*, 961–966.
Leutgeb, S., & Leutgeb, J. K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning & Memory*, *14*, 745–757.
Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*, 1295–1298.
Liu, K. Y., Gould, R. L., Coulson, M. C., Ward, E. V., & Howard, R. J. (2016). Tests of pattern separation and pattern completion in humans-a systematic review. *Hippocampus*, *26*, 705–717.
Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*, 525–551.
Lövdén, M., Schellenbach, M., Grossman-Hutter, B., Krüger, A., & Lindenberger, U. (2005). Environmental topography and postural control demands shape aging-associated decrements in spatial navigation performance. *Psychology and Aging*, *20*, 683–694.
Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal cognitive assessment in community dwelling older adults residing in the southeastern US. *International Journal of Geriatric Psychiatry*, *24*, 197–201.
Marr, D. (1971). Simple memory: A theory for Archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*, 23–81.
McClelland, J. L. (1994). The organization of memory. A parallel distributed processing perspective. *Revue Neurologique (Paris)*, *150*, 570–579.
Molitor, R. J., Ko, P. C., Hussey, E. P., & B a, A. (2014). Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. *Hippocampus*, *24*, 666–672.
Moschner, C., & Baloh, R. W. (1994). Age-related changes in visual tracking. *Journal of Gerontology: Medical Sciences*, *49*, M235-8.
Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., … Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science*, *297*, 211–218.
Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*, 695–699.
Neunuebel, J. P., & Knierim, J. J. (2014). CA3 retrieves coherent representations from degraded input: Direct evidence for CA3 pattern completion and dentate gyrus pattern separation. *Neuron*, *81*, 416–427.
O'Reilly, R. C., Norman, K. A., & McClelland, J. L. (1998). A hippocampal model of recognition memory. In M. I. Jordan, M. J. Kearns, & S. A. Solla (Eds.), *Neural information processing systems* (Vol. 10, pp. 73–79). Cambridge, MA: MIT Press.
O'Shea, A., Cohen, R. A., Porges, E. C., Nissim, N. R., & Woods, A. J. (2016). Cognitive aging and the hippocampus in older adults. *Frontiers in Aging Neuroscience*, *8*, 1–8.
Paleja, M., & Spaniol, J. (2013). Spatial pattern completion deficits in older adults. *Frontiers in Aging Neuroscience*, *5*, 3.
Pertzov, Y., Avidan, G., & Zohary, E. (2009). Accumulation of visual information across multiple fixations. *Journal of Vision*, *9*(10), 1–12.
Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique (Les problèmes). *Archives de Psychologie*, *28*, 215–285.
Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 21159–21164.
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.
Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*, 2442–2449.
Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learning & Memory*, *16*, 338–342.
Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*, 374–391.
Vazdarjanova, A., & Guzowski, J. F. (2004). Differences in hippocampal neuronal population responses to modifications of an environmental context: Evidence for distinct, yet complementary, functions of CA3 and CA1 ensembles. *Journal of Neuroscience*, *24*, 6489–6496.
Vieweg P. 2017. *Recognition memory associated with hippocampal pattern completion in young and older adults*. PhD Thesis, Otto von Guericke University, Magdeburg.
Vieweg, P., Stangl, M., Howard, L. R., & Wolbers, T. (2015). Changes in pattern completion—A key mechanism to explain age-related recognition memory deficits? *Cortex*, *64*, 343–351.
Wechsler, D. (2008). *Wechsler adult intelligence scale—Fourth edition*. San Antonio, TX: The Psychological Corporation.
Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: Prior memories hinder new hippocampal encoding. *Trends in Neurosciences*, *29*, 662–670.
Yassa, M. A., Lacy, J. W., Stark, S. M., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2011). Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*, 968–979.
Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 8873–8878.
Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*, 515–525.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Vieweg P, Riemer M, Berron D, Wolbers T. Memory Image Completion: Establishing a task to behaviorally assess pattern completion in humans. *Hippocampus*. 2019;29:340–351. https://doi.org/10.1002/hipo.23030