Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–18

DOI: 10.1177/17470218221116610
qjep.sagepub.com
SAGE

Where was the toaster? A systematic investigation of semantic construction in a new virtual episodic memory paradigm

Carina Zöllner[1*], Nicole Klein[1*], Sen Cheng[2], Ricarda I Schubotz[3,4], Nikolai Axmacher[5] and Oliver T Wolf[1]

Abstract

Retrieved memories of past events are often inaccurate. The scenario construction model (SCM) postulates that during encoding, only the gist of an episode is stored in the episodic memory trace and during retrieval, information missing from that trace is constructed from semantic information. The current study aimed to find behavioural evidence for semantic construction in a realistic, yet controlled setting by introducing a completely new paradigm and adjusted memory tests that measure semantic construction. Using a desktop virtual reality (VR), participants navigated through a flat in which some household objects appeared in unexpected rooms, creating conflicts between the experienced episode and semantic expectations. The manipulation of congruence enabled us to identify influences from semantic information in cases of episodic memory failure during recall. Besides, we controlled for objects to be task-relevant or task-irrelevant to the sequence of action. In addition to an established old/new recognition task we introduced spatial and temporal recall measures as possible superior memory measures quantifying semantic construction. The recognition task and the spatial recall revealed that both congruence and task-relevance predicted correct episodic memory retrieval. In cases of episodic memory failure, semantic construction was more likely than guessing and occurred more frequently for task-irrelevant objects. In the temporal recall object-pairs belonging to the same semantic room-category were temporally clustered together compared with object-pairs from different semantic categories (at the second retrieval delay). Taken together, our findings support the predictions of the SCM. The new VR paradigm, including the new memory measures appears to be a promising tool for investigating semantic construction.

Keywords

False memory; memory errors; generative episodic memory; memory trace; semantic information; prior knowledge; semantic construction; virtual reality; memory retrieval

Received: 17 November 2021; revised: 7 July 2022; accepted: 8 July 2022

Introduction

Episodic memory enables humans to remember a specific episode from their personal past by mentally travelling through subjective time (Tulving, 2002). Imagine you are in urgent need of a birthday present for a friend and think of buying a new seasonal food calendar. To find out whether your friend already owns one, you could try to remember it from the last time you were in your friend's kitchen. During mental time travel, scenarios of a previously experienced episode can be (re-) constructed. But how accurate are those scenarios? Empirical research suggests that episodic memory traces store only the gist, not the details, of an episode (Bartlett, 1932; Deese, 1959; Gernsbacher, 1985; Koutstaal & Schacter, 1997; Roediger & McDermott, 1995). Semantic information is hypothesised to aid in scenario

[1]Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany
[2]Department of Computational Neuroscience, Institute for Neural Computation, Faculty of Informatics, Ruhr University Bochum, Bochum, Germany
[3]Department of Biological Psychology, Faculty of Psychology, University of Muenster, Muenster, Germany
[4]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany
[5]Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum, Germany

*Carina Zöllner and Nicole Klein contributed equally to this publication.

Corresponding author:
Oliver T Wolf, Department of Cognitive Psychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44780 Bochum, Germany.
Email: oliver.t.wolf@rub.de

construction. In the scenario construction model (SCM), Cheng et al. (2016) support the notion of remembering the gist and add the proposition that all required information missing in this trace is substituted from semantic memory. Consequently, retrieval is thought to be substantially influenced by pre-existing knowledge and expectations. The SCM is supported by previous studies that showed that episodic memory recall is error-prone, and that semantic information influences encoding and retrieval of episodic memory (Chadwick et al., 2016; Roediger & McDermott, 1995). For example, a semantically congruent context during encoding seems to enhance memory recall (Brod & Shing, 2019; Staresina et al., 2009). Revisiting the above example, when you think of your friend's kitchen, you may remember that you had tea during your last visit (the gist of the episode). Nevertheless, it is unlikely that you recall all details of the episode correctly, like the exact colour of the teacup, or whether a calendar was already hanging on the wall. The lack of comprehensive encoding might result in false memories of that episode, for example, that the teacup was white or that a calendar was hanging on the wall—when in fact this information was substituted from your semantic knowledge (e.g., see Brod & Shing, 2019).

We are interested in examining whether these false memories are systematically influenced by interference with the semantic memory system or emerge randomly. Thus, to experimentally investigate the theoretical framework on semantic construction during retrieval, it is important to distinguish between the information stored in the episodic memory trace, and semantic information from the beginning onwards. Hence, during the encoding of an episode, we placed objects congruent to their semantic expectation or at odds. The congruence manipulation enables us to create conflicts between episodic memory traces and semantic knowledge during encoding. Subsequently, in a later memory test, the accuracy of the recalled location of the schematically unexpected (incongruent objects) can give us information about the predominant involvement of the episodic memory trace or semantic information during scenario construction. Nevertheless, congruence can influence memory in two possible ways, already at encoding. On the one hand it may be plausible that incongruent objects are more likely to be remembered better than congruent objects due to the unexpectedness of their location. Küppers and Bayen (2014), for example, found that participants show better source memory for schematically unexpected compared with schematically expected items. On the other hand, the congruence effect describes the finding that participants show better memory performance for items that were presented congruently to their semantically fitting context (Bein et al., 2015). In an experimental study, Brewer and Treyens (1981) placed participants under a pretext in a realistic office. Inside this office, objects were placed that were either schema congruent or schema incongruent. A subsequent free recall test provided evidence for high correlation between schema expectation and recall of objects. Furthermore, participants remembered highly schema-relevant objects that in fact were not present. The later mentioned congruence effect goes in line with the SCM and thus will be part of our hypotheses. Importantly, we wanted to point out again that the congruence manipulation during the encoding phase allows us to test the central prediction of the SCM looking at the ultimate recalled information.

Another factor manipulated at encoding that facilitates investigating semantic construction, is task-relevance. Participants had to interact with half of the objects that were task-relevant. Furthermore, interacting with objects makes the induced episode more realistic and it takes current findings of memory and task-relevance into account as well. A study from Williams et al. (2005), for example, showed that objects that are relevant to the task are remembered better. Taking the factor task-relevance into account helps us to get an idea how strengthening (by task-relevance) or weakening (in terms of task-irrelevance) the episodic memory trace could result in lower or higher semantic construction, respectively. Thus, it could provide us with stronger directions for future hypotheses and alternative manipulations of systematically affecting memory traces.

To systematically investigate semantic construction by manipulating congruence while controlling for possibly influencing factors like task-relevance, we need a complex yet controlled setting. Virtual reality (VR) paradigms provide the compromise between experimental control and ecological validity (Hardiess et al., 2015). A VR system allows to actively navigate by use of input tools (e.g., keyboards). This can be considered as a form of enactment and enhance spatial (Brooks et al., 1999) and episodic (Jebara et al., 2014) memory performance. Enhanced episodic memory when using active navigation in a VR compared with observing passive actions, for example, using video sequences has been shown in previous research (James et al., 2002; Pacheco et al., 2021; Sauzéon et al., 2011).

For the present study, we thus used a desktop VR. Participants experienced an episode by actively navigating through a realistic virtual environment—a three-room flat—and at the same time fulfilling 12 tasks with different household objects in accordance with a cover story. To investigate the influence of semantic information on episodic memory during scenario construction, we made use of a randomised congruence-manipulation by placing objects in the virtual flat: objects appeared in expected, congruent rooms (e.g., a toaster in the kitchen), or in unexpected, incongruent rooms (e.g., a toaster in the bathroom). With this manipulation, we created conflicts between the episodic memory trace and semantic pre-existing knowledge. Recall of the location of

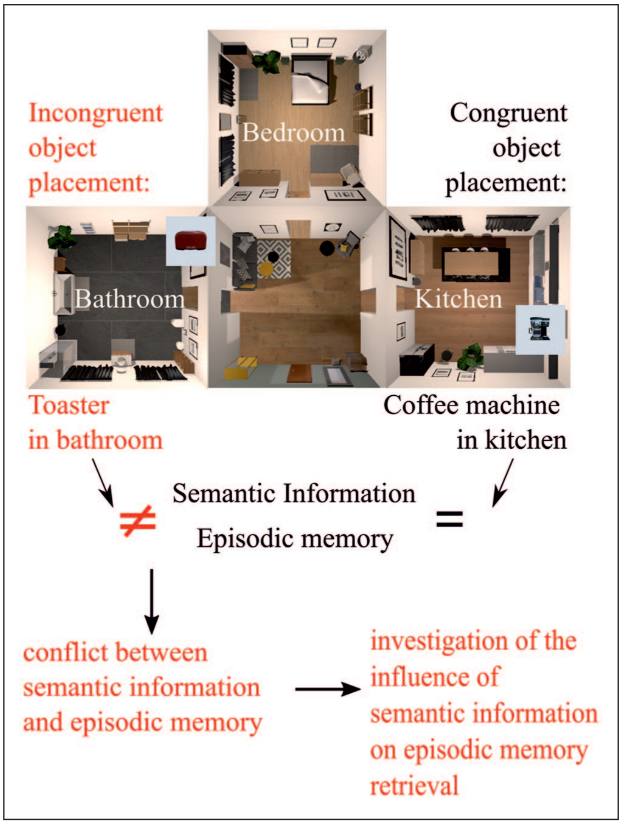

Figure 1. Experimental manipulation in the current study. The creation of conflict between semantic information and episodic memory allows the investigation of the influence of semantic information on episodic recall.

incongruent objects hints at the predominant involvement of one of the components that influences the behavioural response (see Figure 1). For example, if an unexpected location (the toaster in the bathroom) is correctly recalled during retrieval, the response indicates correct retrieval of the episode, while an erroneous memory of it (e.g., the toaster in the kitchen) indicates semantic substitution.

To detect memory errors, we need precise memory measurements. Episodic memory is multimodal, integrating multiple aspects of a past experiences, namely "what happened?" (I've made a sandwich using a toaster), "where" (in the bathroom) and "when" (before. . .). Note that in the following, we will only refer to the where- and when-component, instead of all three, as in our case the what component is conceptually hard to disentangle from the where. The 'what'-component reflects item memory. Nevertheless, in our case the item memory is integrated into the where-component as the main manipulation of the objects is the location (congruence).

Besides, there is the dimension of the temporal relation between whatever happened ("when did it happen?") (Tulving & Donaldson, 1972). Earlier studies showed that when freely recalling a sequence of encoded words, participants tended to cluster semantically related words, which was termed semantic clustering (Bousfield, 1953; Manning & Kahana, 2012). We assume that memory errors will be displayed both in the "where"-component of our paradigm (i.e., in which room incongruent objects were experienced) and in the "when"-component (i.e., interactions with task-relevant objects that semantically belong to the same category [a toaster and a coffee machine] are remembered to have taken place subsequently). Thus, we developed memory measures that capture these components and reflecting the spatiotemporal context.

The commonly used memory measures are the free recall and recognition task. Both held their advantages and disadvantages (for review see Cheke & Clayton, 2013). Therefore, we decided to implement two additional memory measurements that might provide a more specific tool to depict the relevant components for semantic construction, that is, a temporal and spatial recall task (for more details see "Method and material" section). Aiming to find the best memory measure suitable for depicting semantic construction and taking the multiple recall dimensions (of memory) into account, we came up with a combination of the following four memory measures: a free recall (for methodological reasons), a recognition task (for comparison purposes), a spatial recall (for capturing the "where"-component) and a temporal recall (for capturing the "when"-component).

The purpose of our study was twofold. First and foremost, we aimed to test the SCM on a behavioural level. Second, we want to confirm that our paradigm provides a suitable basis for future studies investigating semantic construction. Therefore, we tested the following hypotheses:

1. Participants show better memory performance for objects that were presented congruently to their semantically fitting context (congruence effect).
2. Task-relevant objects are more likely to be recalled episodically correctly, whereas the recall of task-irrelevant objects is more prone to memory errors and consequently more affected by semantic construction.

Only the recall of incongruently placed objects provides us with information about the potential influence of semantic information. Thus, to further investigate scenario construction, in a next step we focus only on the recall of incongruently placed objects.

3. Incongruently placed objects that are recalled episodically wrong are more likely to be erroneously recalled in a semantically congruent location, rather than erroneously recalled in an unrelated location.
4. Incongruently placed and (at the same time) task-relevant objects are more likely to be recalled episodically correctly, whereas incongruently placed and (at the same time) task-irrelevant objects show increased semantic construction.

The temporal recall of only task-relevant objects provides us with additional information about scenario construction, while taking the multidimensionality of memory into account.

5. In the temporal recall, objects belonging to the same semantic category are clustered closer together than objects from a different semantic category, reflecting semantic construction.

We developed a novel VR-based memory paradigm as a promising compromise between ecological validity and experimental control to investigate the interaction between episodic memory traces and semantic information during scenario construction. Besides, we aim to provide the most accurate memory measure capturing scenario construction that takes the multiple memory recall dimensions into account at the same time. Thus, in addition to the commonly used (1) free recall and (2) recognition memory, we introduced two novel developed memory tests for (3) spatial recall and (4) temporal recall. All memory tests are conducted at two different retrieval delays (1 day and 1 week after encoding).

Method and material

Participants

The required sample size was determined a priori using G*Power 3.1 (Faul et al., 2007). Due to a lack of literature on scenario construction, we assumed to find a medium-sized effect for comparing proportions of objects sorted to their semantic category and the proportion of objects sorted to an unrelated room, among all incongruent objects. In this study the within-subject analysis is not independent, thus we will use a Wilcoxon-signed rank test for matched pairs. Accounting for a likely violation of normality, we chose a restrictive estimation of a medium effect size of $dz = .05$ (J. Cohen, 1988), an alpha error probability of .05, and a power of 95%, which revealed a required sample size of 50 participants. To compromise for possible dropout, we recruited 60 participants.

Dropout due to technical problems ($n = 1$), acute motion sickness ($n = 1$), or participants that did not show up at all 3 testing days ($n = 7$) resulted in a total sample size of 51 ($M_{age} = 23$, range = 19–35 years, 32 women, 19 men). Inclusion criteria were (1) general good health, no current or past mental or neurological illness, (2) right handedness, (3) normal or corrected to normal vision, (4) good German language skills, and (5) no self-experienced history of motion sickness. Due to testing during the global COVID-19 pandemic, we took extra precautions, which influenced the screening and hygiene procedure in the experimenter–participant interaction. All participants were naïve concerning the purpose of this study. Participants were either paid 50€ or received course credit.

Software

Virtual environment. We created a virtual flat with the 3D game engine Unity (Unity Technologies, San Francisco) and filled it with realistic details, such as interior and household objects, from several packages (asset store products). The virtual flat consisted of three target rooms (a kitchen, a bathroom, a bedroom) and a neutral entrance room which connects all rooms (see Figure 2).

Materials

All tasks and questionnaires were provided in the German language. All questionnaires (except the demographic questionnaire), the recognition memory task, and the PVT were programmed in MATLAB (2020), using the Psychtoolbox-based OTBR Toolbox (Brainard, 1997; Rose et al., 2008). We used the game engine Unity 3D (version 2020.1.0a24, Unity Technologies, San Francisco) to develop the virtual environment and the spatial and temporal recall.

Stimuli. We used 24 virtual 3D household objects in total. We extracted 22 of them from different Unity packages (for a comprehensive list see the online Supplementary Material 1). In addition, we created two virtual objects ourselves using the 3D computer graphics software toolset Blender (Blender Online Community, 2018). Using a prior online survey, we verified that independent participants rated the objects we used fitting to the same semantic schemata (e.g., a toaster which belongs in the kitchen), resulting in eight objects per room category (see Figure 3).

Study procedure

The study included 3 testing days (see Figure 2). The first 2 days took place consecutively and the last testing day 1 week later. At day 1, episodic memory was encoded by inducing an episode of a preparation phase for a pretended date inside a virtual environment. For methodological considerations of a future study, we distinguished between two conditions, in which half of the participants accomplished an additional task, at day 1. Thus, the testing on day 1 lasted between 40 and 80 min depending on the condition. On day 2 and 8, we tested memory retrieval using four different memory measure tests. The second day lasted about 80 min and the third day lasted about 40 min again. The study was approved by the ethical committee of the Faculty of Psychology at Ruhr University Bochum, Germany.

Day 1. After giving written informed consent, participants answered two questionnaires—one demographic and one comprising the Edinburgh handedness Inventory (EHI; Oldfield, 1971). Half of the participants completed a picture viewing task (PVT).

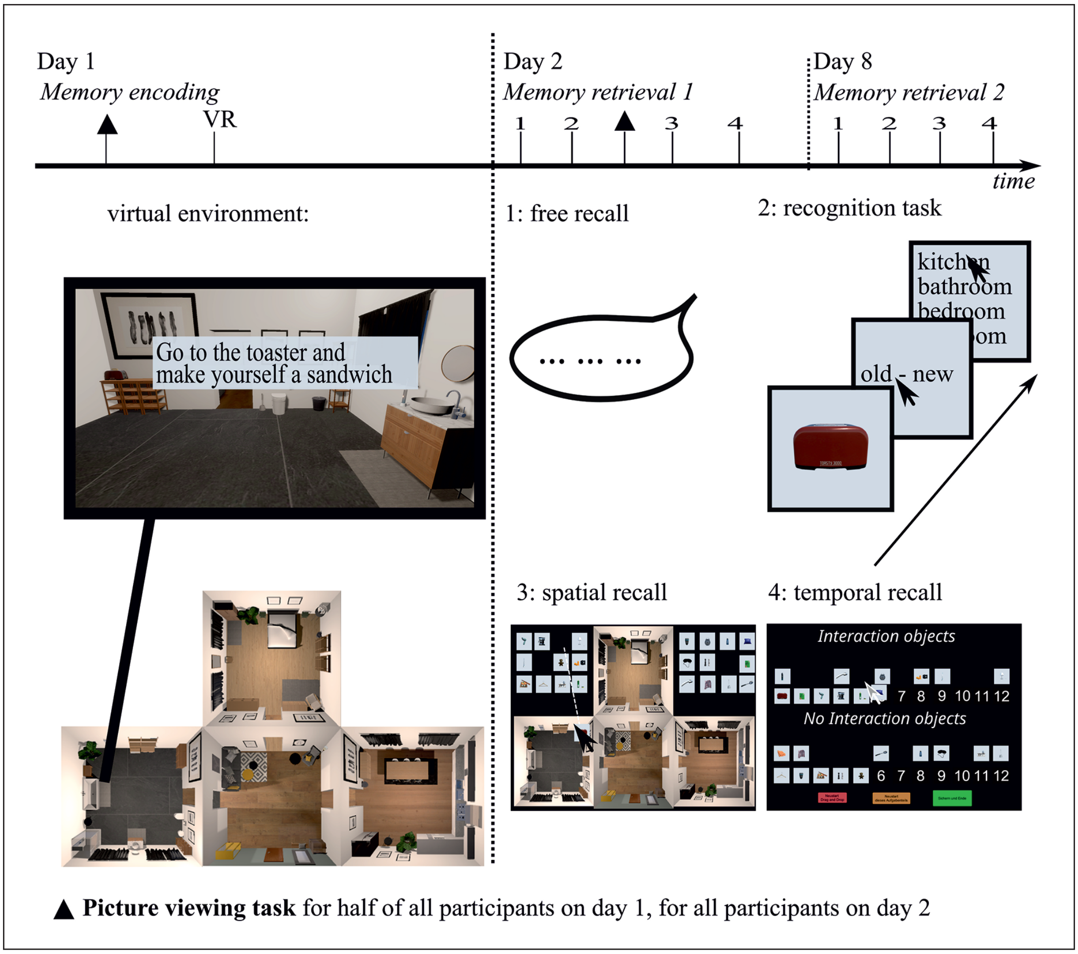

Figure 2. Overview of procedure of all 3 testing days. The first day included the memory encoding in the virtual environment. The other 2 days consisted of four memory retrieval tasks.

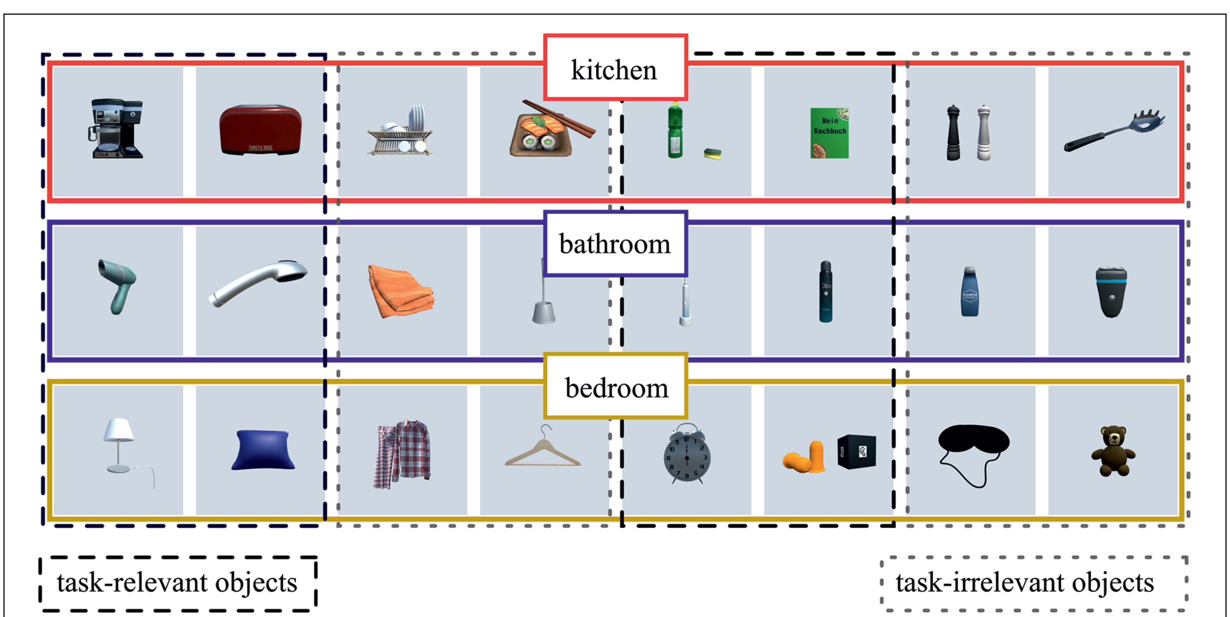

Figure 3. Overview of stimuli.
We used 24 household objects. Red, blue, and yellow frames indicate the room categories of the objects (kitchen, bathroom, and bedroom). black dotted frames mark objects that are task-relevant. Grey dotted frames mark objects that are task-irrelevant.

During the picture viewing blocks, participants saw 2D pictures of 24 household objects in random order on a greyscale background. The pictures were screenshots from virtual 3D objects that are used in the following episodic virtual environment (EVE) task. Each picture was presented for 3,000 ms. Between two stimuli, participants saw a black fixation cross, which changed its colour to red in 5% of the cases. Participants pressed a button whenever they spotted the colour change and watched the presented stimuli attentively. Overall, 10 picture viewing blocks were presented. Between the picture viewing blocks, participants saw 24 pictures of faces in random order during the pause blocks. The faces were randomly drawn from a set of 200 faces of an intern, systematic Google picture search. We chose pictures with a broad variation in diversity, age, and origin. Half of the participants (PVT group) performed this task twice (pre and post encoding, see Figure 2), while the other half (no-PVT group) only performed this task once (post encoding). The PVT was conducted in preparation for a future fMRI version of the paradigm, which will aim to investigate task-induced changes in neural similarity structure by using representational similarity analysis (RSA). For more detailed information on RSA (see Bierbrauer et al., 2021). The PVT is of no further interest for this study. Nevertheless, we will analyse the group differences of PVT and no-PVT group (group membership) in terms of memory differences in the "Results" section. Group membership refers only to the randomised assignment of PVT or no-PVT prior to encoding.

Subsequently, all participants underwent episodic memory encoding in the form of EVE task. During a familiarisation phase with the virtual environment participants navigated freely in first-person perspective (using a keyboard) through a flat. The flat was created with a kitchen, a bathroom, a bedroom, and a hallway connecting all three rooms. Each room was already set up with room-specific interior. Besides, four specific household objects were placed randomly across the flat. A presentation of a cover story (see the online Supplementary Material 2) of recently moving in and doing maintenance work followed. Participants finished two tasks related to the cover story, by approaching the two task-relevant objects and interacting with them (using button presses). To assure that all participants explored the flat in a comparable manner, a 5-min time limit was set. During the exploration phase, participants filled out a questionnaire regarding the flat to ensure that they recognised the rooms and the floor plan. The content of the familiarisation phase was not tested during the later memory recall.

In the following main EVE task, participants received a new cover story of preparing their flat for an evening date. The flat and the interior stayed the same, while instead of the 4 example objects, 24 new household objects (8 objects per room category) were located inside the flat. We manipulated the location of all objects inside the VR regarding congruence to distinguish between episodic end semantic memory recall on the later run. Thus, we placed half of the objects according to their semantically fitting category (e.g., a coffee machine in the kitchen) and the others at odds (e.g., a toaster in the bathroom). In addition, we predefined 12 objects as task-relevant or task-irrelevant, such that task-relevant objects were included in the sequence of actions during the EVE task. This functions to enhance the plausibility of the cover story and to control for the possible influencing factor of attention due to interaction with some objects but not with others, which reflects an everyday life behaviour. All task-relevant and task-irrelevant objects were the same for all participants. Nevertheless, the order of the tasks was randomised between participants. Altogether, participants conducted 12 tasks (for full list see Supplementary Material 3) which were related to the new cover story—for example, "You still need to buy groceries. Look for the recipe inside the cooking book and take a picture of the list of ingredients," with half of the objects. We randomised both congruence and task relevance resulting in the location of eight objects per room—4 congruent (2-task-relevant or-irrelevant) and 4 incongruent objects (2 task-relevant or-irrelevant).

Prior and directly after the EVE task, participants answered the Positive and Negative Affect Schedule (PANAS; Breyer & Bluemke, 2016; Watson et al., 1988) to check for possible influence of the task on participant's affect. After participants answered the Igroup Presence Questionnaire (IPQ; Schubert et al., 2001) assessing subjective feeling of experienced presence in a VR, the first day ended.

Day 2. All participants underwent four incidental memory tasks and the PVT. Starting with the free recall, participants were instructed to freely narrate everything they remember about the episode in the virtual environment, while being audio-recorded. By prior presentation of an exemplary audio recording (referring to the familiarisation phase) we ensured that participants would focus especially on the objects and their location inside the virtual environment during their own recall and thus providing us with the specific details necessary for investigating scenario construction.

The following recognition task consisted of 48 trials. In each trial, a 2D picture of a household object was presented on a plain grey background. In half of the trials "old" objects (i.e., objects that had been present in the virtual flat) were shown, while in the other half new objects were shown, which had not been present in the virtual flat. These "lures" were chosen according to their semantic congruence (eight new objects per room category, i.e., $n=24$) as well as their conceptual similarity to some of the objects (for full list see Supplementary Material 2). Each picture was displayed for a fixed duration of 3 s. The order of presented pictures was randomised over all participants. Per trial participants had 15 s to respond via mouse click on a six-point confidence scale to the question: "Have you seen this object in the flat?" (from "sure new" to "sure old"). When a participant indicated to have seen the object in the flat, the additional question: "In which room have you seen the object" was shown. A new trial started whenever a participant answered this question or rated a presented object as "new." In case no response was given for 15 s, the trial was excluded from the analysis and the next trial started automatically.

After the first two memory measures in which it was crucial that participants relied on their own episodic memory without any cues regarding the actual presence of the objects during the episode, the PVT followed (for details see section "Day 1"). Due to the nature of the following self-developed memory measures, all objects that were present during the EVE task appeared all together to arrange them according to the "where" and "when"-component of memory. Thus, the PVT task did not give away any additional information about the objects that were present during the EVE task, that participants would not receive during the last memory measures, anyway.

In the subsequent self-developed spatial recall, participants saw the 2D floor plan of the virtual flat from a bird's-eye view and 2D pictures of the 24 household objects on a grey background arranged randomly beside it. The task was to drag and drop (using the computer mouse) one object after the other to the specific location inside the virtual flat where participants remembered to have encountered this object during the EVE task. The instructions encouraged them to drag all objects and to guess when they could not remember an item. After dragging an object onto a location and releasing the mouse key the object disappeared from the display. Participants did not receive any feedback about the correctness of their object placement. Participants could restart the trial via a button press if they believed they had chosen a wrong location. To familiarise themselves with the procedure, participants conducted a test trial, including the objects from the familiarisation phase of the EVE task and were given the opportunity to ask questions. For our analysis, we measured memory accuracy as the difference between the distance of remembered and actual position (see Figure 4). The distance from the dropped position to the correct position was calculated considering the direct virtual route between the two points. The diameter of one room (corner to corner) was 290 units. The neutral distance (e.g., kitchen door to bathroom door) was 222 units. Thus, the distance from the dropped position to the correct position equals the sum of the distance from the dropped position to the room door, the neutral distance, and the distance from the target room door to the correct position.

Finally, in our second self-developed temporal recall, participants saw a black screen, with 2D pictures of all task-relevant objects in the order they choose to drag and drop the objects in the previous spatial task. Here they had the chance to focus on the actual order they remember to have performed the task and arrange the pictures of the objects accordingly onto a list of numbers ranging from 1 to 12. Whereas the number one indicated the first task and number 12 the last task, respectively. Participants were able to rearrange the order in the virtual space by using a computer mouse. The task is depicted in Figure 2. Technical failures in this task specifically led to a final sample size of $N=46$ participants for the following analyses. Then, participants sorted pictures from all 24 household objects, onto a scale from 1 to 24, regardless of their task-relevance but concerning the order participants remember to have noticed them in the VR. After the last memory test, day 2 ended.

Day 8. Participants underwent all four memory tasks again. The experiment ended with the final survey of autobiographical memory (SAM; Palombo et al., 2013), which assesses general memory.

Statistical methods

Data preparation. We prepared the data for further analyses with Python 3.7 implementation in *Spyder* (Raybaut, 2009; van Rossum & Drake, 2009). All statistical analyses were conducted with *R* implementation in *RStudio* (R Core Team, 2019; RStudio Team, 2019). In a first step we had a general look at our four memory measures: (1) free recall, (2) recognition memory (old/new ratings and subsequent room-sorting), (3) spatial memory (the distance between the drop location and the correct location of an object in the virtual environment, depicted in Figure 4a, in unity units, and according room-sorting), and (4) temporal recall (a recall of the order of the tasks). The descriptive analysis of the data from the free recall task revealed that only 37% of the objects were mentioned at all in the free recall, and for only 32% an according room was mentioned. We acknowledged that the amount of data acquired from this recall task—on average only seven object-room-recalls per participant—is insufficient for our planned statistical contrasts (i.e., estimating the effect of congruence on memory recall or differentiating between different room-sortings for only incongruent objects and estimating the effect of task-relevance). We gained a first insight into our data by analysing the recognition memory task. To this end, we first calculated and reported general recognition memory performance with the sensitivity measure d' (estimated from the hit and false alarm rate) and then, including only targets but not lures, we further estimated a model predicting correct "old"-responses (hits) by an objects' congruence, "task-relevance" and the time point of retrieval. Regarding the room-sorting, however, we looked at the relatedness between the room responses in the

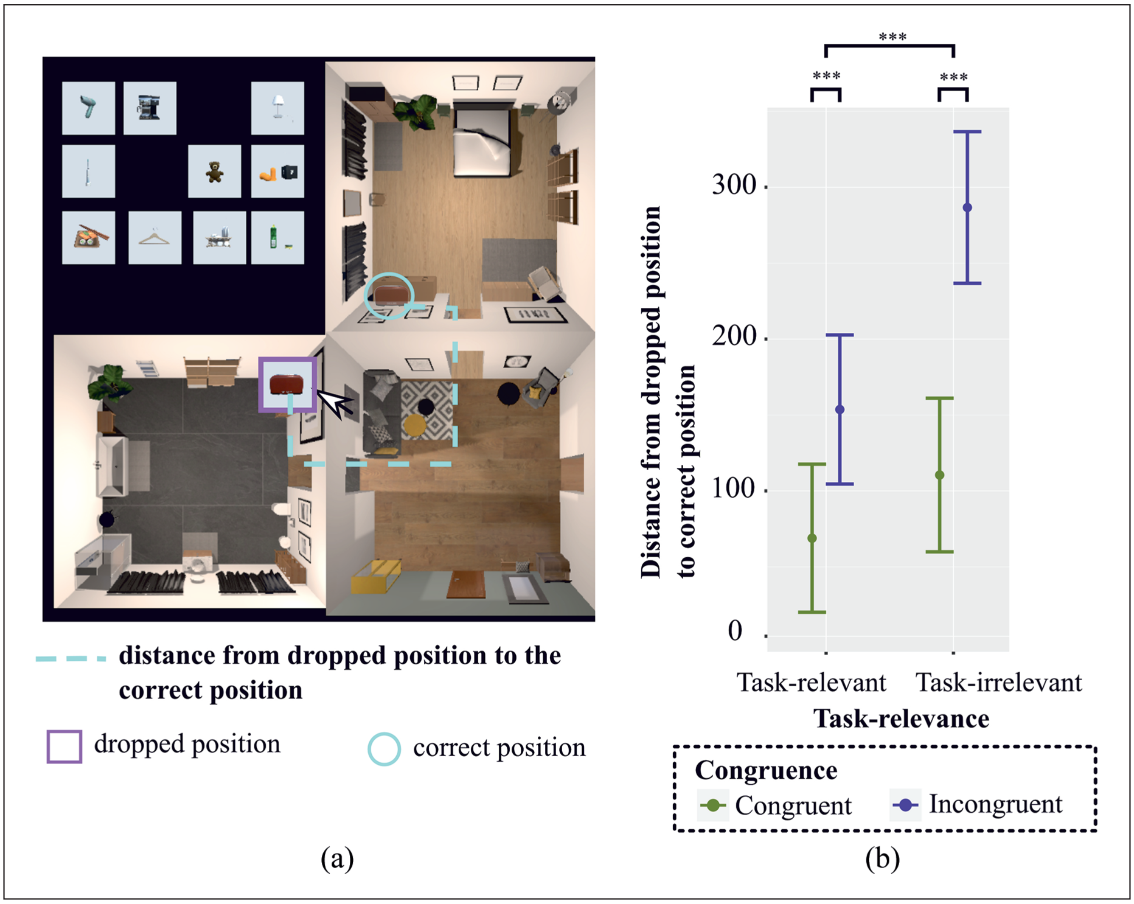

Figure 4. Accuracy of spatial memory is modulated by congruence and task-relevance. (a) In this example, the distance from the dropped position to the correct position equals the sum of the distance from the dropped position to the bathroom door, the neutral distance, and the distance from the bedroom door to the correct position. (b) "Congruence" and "task-relevance" significantly predict the distance from the dropped to the correct position for all initially recognised objects across delays, groups, and sizes.

Depicted are the model estimates and standard errors. Post hoc analyses revealed that congruence predicted memory for both task-relevant and task-irrelevant objects, and that the difference is higher for task-irrelevant objects.

$^{***}p < .001$.

recognition task and the participant's chosen rooms in the spatial recall task using Pearson correlation analyses with the R-package *psych* (Revelle, 2020). We expected that both tasks are not orthogonal to each other, because the two tasks assess similar content (i.e., which room the participant remembered to have seen a specific object). As predicted, the room-sorting performance in the recognition memory task was highly correlated with the room-sorting in the spatial recall task ($r = -.74$). Thus, in our investigation of the main hypotheses, we restricted our analyses to the room-sortings from the spatial recall task. To preclude guessing, we furthermore excluded objects that were not recognised as "old" in the recognition task, and thus possibly not noticed at all, from these analyses.

Data analysis. We started by generally analysing how relevant object characteristics ("congruence" and "task-relevance") and how the within-subject factor "retrieval delay" and the between-subject factor "group-membership" affected memory-performance. We first analysed whether the hit rate (was a target object correctly identified as "old") differed between the two object-characteristics ("congruence" and "task-relevance"), the two retrieval-delays and the two groups, including interaction-effects between "congruence," "task-relevance" and "group-membership." We conducted a logistic linear mixed model analysis, taking both individual subject effects and object effects into account by including them as random factors in our model, as an inclusion of both factors improved the model fit (estimated with the intraclass correlation coefficient [ICC] subject ICC = .06, object ICC = .29). We followed the procedure following Sommet and Morselli (2017), and thus estimated significance by computing 95% confidence intervals (CI) and interpreting odds ratio (OR, i.e., if the value 1 is part of the 95% CI, there is no significant effect of that respective predictor). We centred all predictors to mean 0.

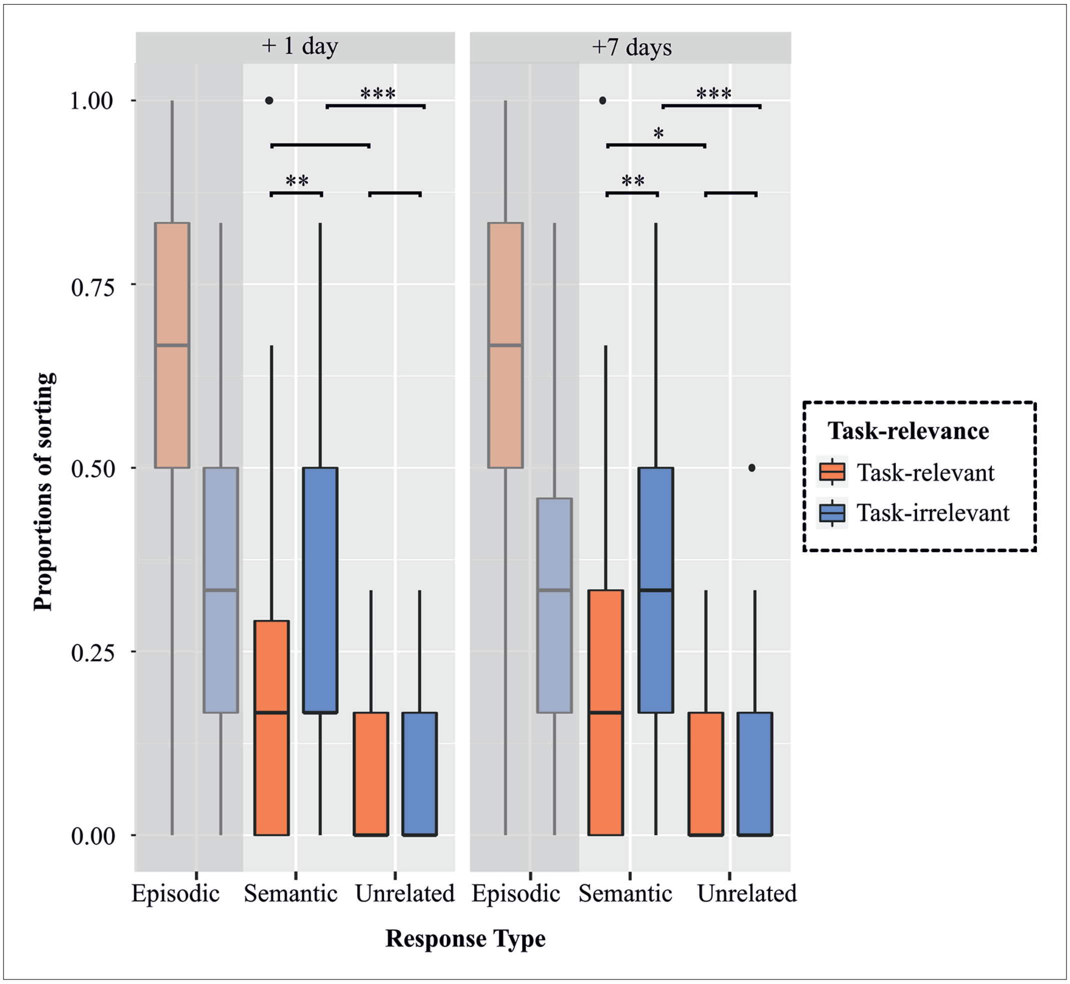

Figure 5. During the spatial recall task, all 12 incongruent objects were either (a) sorted episodically correctly, (b) sorted semantically, (c) sorted to the unrelated room, or (d) excluded, because they were rated as "new" in the recognition task. Depicted are the resulting proportions of room-sortings for all participants, separately for task-relevant and task-irrelevant objects and both retrieval delays. The box-plots show the median, the first and third quartiles (represented by the box), the minimum and maximum (represented by the whiskers), and outliers (data points which are smaller or larger than 1.5 times the interquartile range). Two-sided paired Wilcoxon signed rank sum tests (represented by the connecting, horizontal bars) revealed that on both retrieval delays, task-irrelevant objects were more likely sorted to their semantic category rather than to the unrelated room. This is also the case for task-relevant objects after the second, but not after the first retrieval delay. Furthermore, task-irrelevant objects were more likely sorted to their semantic category than task-relevant ones. Proportions of episodic sortings were not included in statistical analyses and are marked as shaded. A fourth and complementing column ("Proportion of no response," including all objects which were not recognised, or no response was given on) was not visualised for reasons of simplicity.
$^{**}p_{\mathrm{Holm}} < .01$, $^{***}p_{\mathrm{Holm}} < .001$.

In a next step, we conducted a similar analysis for the spatial recall, but included only objects which were correctly identified as "old" in the recognition task. We analysed whether spatial recall performance (measured with the walking-distance between the drop-location and the correct location of an object, "drop error," depicted in Figure 5a) could be predicted by "congruence," "task-relevance," "retrieval delay" and "group-membership," again including interaction-effects between "congruence," "task-relevance" and "group-membership." Prior to this analysis, we again confirmed clustering in our data and could show that allowing for the two random factors "subject" and "object" improved the model fit (ICC = .209). We hence used linear mixed models to account for these random factors, using the R-package *lme4* (Bates et al., 2015). We used restricted maximum likelihood estimation and tested statistical significance of fixed effects with a Type III ANOVA *F*-statistics using the R-package *car* (Fox & Weisberg, 2019). We conducted post hoc pairwise comparisons with *t*-tests and adjusted *p*-values using "Tukey"-correction in the R-package *emmeans* (Lenth, 2020).

Our second research question focuses on the memory for incongruent objects, as they allow for a discrimination between episodic memory (sorting to the correct room) and semantic construction (sorting to the semantically related room). Using the room-sorting data from the

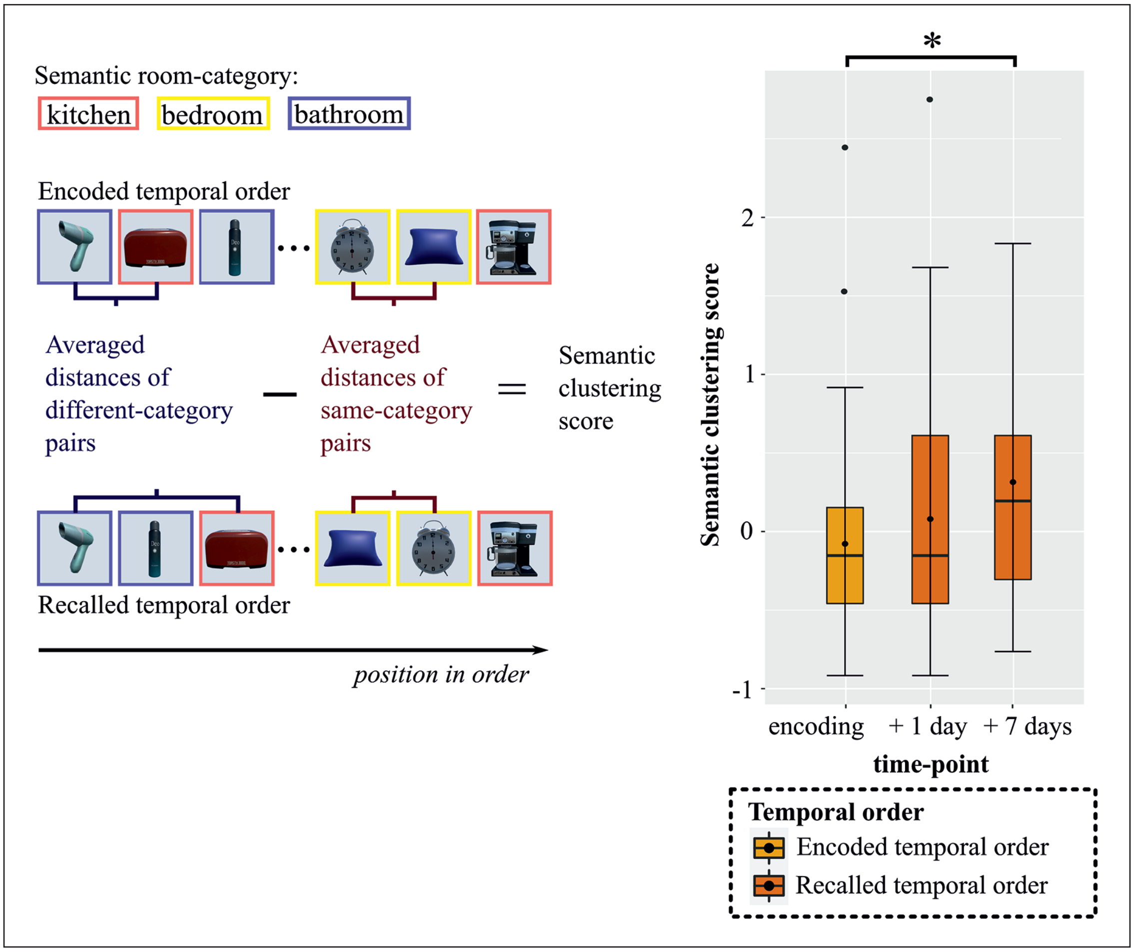

Figure 6. Depicted is the schematic overview of the analysis of the temporal recall data and the plotted results. First, the average distances between object-pairs belonging to different semantic categories and the average distances between object-pairs belonging to the same semantic category were calculated. The difference-value makes the semantic clustering score, which was calculated for the recalled temporal order after both retrieval delays, and the actually encoded temporal order.
The box-plots show the median, the first and third quartiles (represented by the box), the minimum and maximum (represented by the whiskers), and outliers (data points which are smaller or larger than 1.5 times the interquartile range). In addition, the points within the box represent the mean value. The encoded semantic clustering score ($M = -0.078$, $SD = 0.522$) differed from the retrieved semantic clustering after a delay of 7 days ($M = 0.315$, $SD = 0.820$), but not after a delay of 1 day ($M = 0.080$, $SD = 0.775$).
*$p < .05$.

spatial recall task of all 12 incongruently encountered objects, we calculated the proportions of objects sorted to their semantically related room, to the unrelated room, and, for illustrating purposes, also to the correct room. The proportions were calculated separately for task-relevant and task-irrelevant objects. For each of our two retrieval delays, we first investigated whether the proportions differed between task-relevant and task-irrelevant objects. Importantly, we then analysed the semantic bias, that is, differences between the proportion of objects sorted to the semantic room (semantic construction) and the proportion of objects sorted to the unrelated room (guesswork/ chance level), separately for task-relevant and task-irrelevant objects. For statistical comparisons we used two-sided paired Wilcoxon signed rank sum tests. We accounted for multiple testing (8 tests) using Holm correction.

Finally, we analysed whether the temporal recall was influenced by semantic information. We were especially interested in whether the temporal order, in which the task-relevant objects were sorted by the participants, was significantly influenced by the semantic relation between two objects, that is, whether objects semantically belonging to the same room are temporally clustered together as a sign for semantic construction in temporal recall. For both retrieval delays, we estimated the difference value between the averaged recalled temporal distances of object-pairs belonging to the same semantic category and the recalled temporal distance between object-pairs belonging to different semantic categories as a measure of semantic clustering (see Figure 6). In addition, we estimated the same semantic clustering score for the actual, encoded order of the tasks, to make sure the recalled semantic clustering was not simply representing the encoded clustering, despite randomisation the order of the tasks across subjects. Then we compared the recalled semantic clustering of task-relevant objects with the encoded semantic clustering of objects using paired, two-sided t-tests.

Results

General task information

Virtual environment task. On average, participants have spent 6.4 min ($SD=2.6$ min) in the virtual apartment solving the 12 tasks. On average, each task lasted 28.26 s ($SD=33.7$). Participants spent the majority of it searching for the item (time between the first appearance of the instruction and first interaction), which lasted on average 27.00 s ($SD=33.06$). The search time differed between congruent ($M=22.66$, $SD=26.79$) and incongruent ($M=31.60$, $SD=38.11$) objects. The remaining time (task time, $M=1.25$, $SD=2.33$) on average did not differ much between congruent ($M=1.16$, $SD=1.16$) and incongruent ($M=1.35$, $SD=3.12$) objects.

Memory retrieval. After the first retrieval delay, participants freely recalled on average 7.83 objects ($SD=2.66$, second retrieval delay: $M=10.13$, $SD=3.93$). In the recognition memory task, participants showed an average recognition memory performance (d') of 1.51 at the first retrieval delay ($SD=0.64$, second retrieval delay: $M=1.87$, $SD=0.69$), with an average hit rate of 0.77 ($SD=0.1$, second retrieval delay: $M=0.84$, $SD=0.1$) and a false alarm rate of 0.14 ($SD=0.12$, second retrieval delay: $M=0.16$, $SD=0.12$). During the spatial recall task, they dropped all objects on average 191.09 unity units away from their original position ($SD=60.63$, second retrieval delay: $M=194.99$, $SD=65.92$). In the temporal recall, task-relevant objects were sorted on average 3 units away from the original position in the order of the tasks ($SD=0.98$, second retrieval delay: $M=3.1$, $SD=1.04$).

Object memory performance

First, we looked at how the hit rate from the recognition memory task was predicted by object characteristics ("task-relevance" and "congruence" as "within-object" factors), the within-subject factor "retrieval delay," and the between-subject factor "group-membership" (PVT–no-PVT). Generally, we found that an object was more likely recognised as "old" when it was task-relevant as compared with task-irrelevant objects (OR = 0.118, 95% CI [0.064, 0.208]). Furthermore, the likelihood of the correct recognition of a target object was significantly higher at the second retrieval 1 week after encoding, compared with the first retrieval (OR = 1.762, 95% CI [1.414, 2.202]). In our analyses, the PVT-group was more likely to identify an object as "old" as compared with the no-PVT group (OR = 1.623, 95% CI [1.076, 2.469]). Although we did not find a significant effect of whether an object was congruent, or not (OR = 1.318, 95% CI [0.998, 1.760]), we found a significant interaction between "congruence" and "task-relevance" (OR = 0.456, 95% CI [0.256, 0.794]), such that task-relevant objects were more likely identified as "old" when they were congruently encountered ($M=0.95$, $SD=0.23$) rather than incongruently ($M=0.91$, $SD=0.28$), and task-irrelevant objects were more likely identified as "old" when they were incongruently encountered ($M=0.70$, $SD=0.46$) rather than congruently ($M=0.67$, $SD=0.47$). There was no significant interaction between "group-membership" and "congruence" (OR = 1.170, 95% CI [0.669, 2.082]), "task-relevance" (OR = 0.595, 95% CI [0.334, 1.031]) or the two of them (OR = 0.773, 95% CI [0.244, 2.362]).

Then, we investigated the effects of the same predictors on spatial memory accuracy, now excluding all objects not identified as "old" in the recognition task to prevent effects of mere guessing. To this end, we again conducted a linear mixed model, using restricted maximum likelihood estimation and tested statistical significance of fixed effects with a Type III ANOVA F-statistics. The key results are depicted in Figure 4b. We found that congruent objects ($M=88.37$, $SD=143.9$ unity units) were significantly better recalled than incongruent objects, $M=238$, $SD=245.86$, F (1, 1967.28) = 383.685, $p<.001$. Also, objects which were task-relevant ($M=115.68$, $SD=180.4$) were recalled significantly better than task-irrelevant objects, $M=227.58$, $SD=239.35$, F (1, 19.77) = 30.010, $p<.001$. Spatial memory performance was decreased 1 week as compared with 1 day after encoding, F (1, 1962.46) = 4.577, $p=.0325$. We also found a significant interaction between the congruence of an object and its task-relevance, F (1, 1967.27) = 71.164, $p<.001$. Post hoc pairwise comparisons revealed better memory for congruent than incongruent objects of the task-relevant, t (1969.9) = −8.610, $P_{\text{Tukey}}<.001$, and task-irrelevant, t (1985.1) = −18.402, $P_{\text{Tukey}}<.001$, category. As visible in Figure 4b, the difference in memory recall accuracy between congruent and incongruent objects was higher for task-irrelevant objects compared with task-relevant objects. Furthermore, our analysis showed that group-membership (pre-PVT vs. no pre-PVT), did not influence spatial memory, F (1, 52.26) = 0.002, $p=.961$, nor was there an interaction of group with any other predictor, group × congruence: F (1, 1963.50) = .003, $p=.956$, group × task-relevance: F (1, 1962.74) = .742, $p=.389$, group × congruence × task-relevance: F (1, 1963.85) = .039, $p=.844$.

Semantic construction

We further studied the influence of semantic information on recall and focused on the room-sortings of incongruent objects. To this end, we first looked at whether the proportion of objects sorted to their semantic category among all incongruent objects and the proportion of objects sorted to the unrelated room among all incongruent objects per participant on both retrieval delays differed between task-relevant and -irrelevant objects. As the assumption of normality was violated, we used a two-sided paired Wilcoxon signed rank test. On both retrieval delays, task-irrelevant objects were more likely sorted to their semantically fitting category than task-relevant objects (1 day: $V=132, p_{\text{Holm}} < .01$, 7 days: $V=136.5, p_{\text{Holm}} < .01$). However, there is no difference in the proportions of objects sorted to their unrelated room between task-relevant and task-irrelevant objects (1 day: $V=189$, $p_{\text{Holm}}=1$, 7 days: $V=230$, $p_{\text{Holm}}=1.000$). In a next step, we analysed the semantic bias separately for task-relevant and -irrelevant objects using two-sided paired Wilcoxon signed rank test to test for differences between the proportions of semantic sorting (i.e., semantic construction) and the proportions of unrelated sorting (guesswork/chance level). On both retrieval delays, task-irrelevant objects were significantly more likely sorted into the room that matched their semantic category than into the unrelated room (1 day: $V=739.5$, $p_{\text{Holm}} < 0.001$; 7 days: $V=774$, $p_{\text{Holm}} < .001$). For task-relevant objects however, the effect was only present 7 days ($V=352.5$, $p_{\text{Holm}} < .05$), but not 1 day after encoding ($V=342.5, p_{\text{Holm}}=.173$). All results are included in Figure 5.

Semantic construction in temporal recall

We analysed whether there was a difference in the recalled temporal proximity between object-pairs semantically belonging to the same room or to different rooms. To this end, we estimated a temporal semantic clustering score by calculating the difference value between the averaged temporal distance of object-pairs semantically belonging to different rooms and the averaged temporal distance of object-pairs belonging to the same semantic room-category (see Figure 6), for both, the recalled temporal order and encoded temporal order. The paired two-sided *t*-test comparing the two scores indicated that the recalled semantic clustering was higher than the encoded semantic clustering 1 week, $t(47)=2.805$, $p_{\text{Holm}} < .05$, but not 1 day after encoding, $t(43)=1.238$, $p_{\text{Holm}}=0.223$. Thus, in the temporal recall 1 week after encoding, the order of the tasks from objects belonging to the same semantic room-category was remembered to be temporally clustered together, compared with the encoded temporal order.

Replication study

Method and material

To strengthen our results, we conducted a replication study, which only included a subset of the tasks from the main study. This replication study aimed to replicate the previous results with the EVE task during encoding and the spatial recall task during retrieval 1 day after encoding. Thereby, we excluded any influences on the retrieval from the PVT or the other retrieval measures applied in the main study. We conducted data collection during a bachelors' degree seminar. We initially tested 52 participants to mirror the previous data collection. However, eight participants had to be excluded due to technical failure or non-compliance.

Participants first underwent the EVE task, which was improved by balancing task-relevant and task-irrelevant objects (i.e., each participants encountered a different subset of half of all 24 objects as task-relevant, still performing four tasks per room, two of which were congruent objects and two incongruent ones). One day after encoding, participants underwent an online version of the spatial recall task at home on a computer or laptop, in which they dragged and dropped every object to the place they remember to have encountered it. Online-recall was conducted using the jsPsych (de Leeuw, 2015) implementing platform cognition.run. For the purpose of the seminar, an additional group manipulation was included (half of the participants received a written, positive feedback after the fulfilment of each task, i.e., "Wow."), and additional questionnaires were presented to the participants. The data were prepared similarly to the main study, using the Python implementation in Spyder. Using linear mixed models, we first analysed whether object characteristics as congruence and task-relevance predicted the drop-error. We confirmed clustering in the data and setting participants, but not objects, as random factors improved the model fit (ICC = .04). In a second step, we estimated the rooms in which objects were placed. Similar to the main study, including only incongruently encountered items, we then calculated the proportions of incongruent objects dropped in the correct/episodic room, semantically related room or wrong room for every participant. We statistically analysed whether objects were more likely placed in the semantically fitting room rather than in the wrong, unrelated room in cases of incorrect memory using paired, two-sided Wilcoxon signed rank sum tests.

Results

Replicating the results from the larger study, we found a significant main effect for congruence and task-relevance predicting the drop-error. Congruent objects were placed significantly closer to their original position than incongruent objects, $F\ (1, 958.53)=169.654$, $p < .001$, and task-relevant objects were placed significantly closer to their original position than task-irrelevantt objects, $F\ (1, 958.52)=119.022$, $p < .001$. Furthermore, we found a significant interaction between congruence and task-relevance, $F\ (1, 958.48)=36.658$, $p < .001$. The results match the results from the old study, as post hoc comparisons also revealed better memory for congruent than incongruent objects among the task-relevant objects, $t\ (959)=-4.941$, $p_{\text{Tukey}} < .001$, and task-irrelevant objects, $t\ (959)=-13.463$, $p_{\text{Tukey}} < .001$. Again, the difference in drop-error between congruent and incongruent objects is higher for task-irrelevant than task-relevant objects. The results are depicted in Figure 7a.

We could additionally replicate the findings regarding semantic construction. In the case of episodic memory failure, that is, when participants failed to recall the correct room in which they actually encountered the object, we found that 1 day after encoding, task-irrelevant objects were more likely placed in the semantically fitting room rather than in the wrong, unrelated room ($V=750.5$, $p_{\text{Holm}} < .001$). This was not true for task-relevant objects ($V=193$, $p_{\text{Holm}} = .809$). Relatedly, the proportion of objects sorted to the semantically fitting room was significantly higher for task-irrelevant objects as compared with task-relevant objects ($V=22.5$, $p_{\text{Holm}} < .001$). There was no such difference for the proportions of objects sorted to the unrelated room ($V=126.5$, $p_{\text{Holm}} = .809$). The results are depicted in Figure 7b.

Discussion

The main goal of the current study was to experimentally test the hypothesis that semantic information enriches episodic memory retrieval by developing an ecologically valid virtual navigation task. By placing a subset of objects at incongruent spatial locations, we were able to create conflicts between episodic memory and semantic expectations. We found that congruence and task-relevance increased the likelihood to correctly recall an object. If participants correctly recognised an object they had encountered in the virtual flat, but placed it in an incorrect room, we observed that they recalled an object's location significantly more often based on semantic expectations (semantic construction) rather than guessing. The effect occurred during both retrieval delays and was significantly more pronounced for task-irrelevant objects. This indicates that the likelihood of semantic construction is higher in cases of weak episodic memory traces and supports the predictions of the SCM (Cheng et al., 2016). As expected, recall accuracy of episodic memory was high whenever there was a semantic match between the actual location and the object's typical semantic location.

Our overall finding that memory performance was substantially influenced by semantic information, both in our spatial and in our temporal recall measures, is in line with previous research. Koutstaal and Schacter (1997) showed, for example, that the false recognition of lures was more likely if lures were from the same category as previously studied exemplars than if they were from a different category. The same was shown by the classic studies of Bartlett (1932), who reported that memory recall of stories was influenced by expectations and cultural background. Recently, Sipe and Pathman (2020) had children perform location-congruent or location-incongruent tasks in different physical buildings. Not only was the memory accuracy for congruent tasks higher, they also showed that in the case of memory error, a recall of the location of the task according to its semantically fitting category instead of the actual location was higher than chance, which is consistent with our findings. For their experiment, the authors modified a real-world, local museum, so that children could physically walk around for encoding. The fact that we found comparable results in our design illustrates that the improvement in economic aspects by using a VR is not at the cost of ecological validity.

The influence of semantic information on memory has been widely discussed and could happen during encoding and during retrieval. For example, Tulving (2001) proposed in his serial parallel independent (SPI) model that during encoding, information is processed in interdependent stages, that is, first by the perceptual system, then by the semantic system, and finally by the episodic system.

Semantic information could, however, also directly influence memory retrieval, as suggested by Cheng and colleagues (2016) in their SCM and, in principle, also implied by Anderson (1983), who proposed that memory is stored as an interconnected network of units of information, and that during retrieval, not single bits of information but whole networks are retrieved. Thus, recalling a past episode is likely accompanied by the activation of related memory traces, which leads to semantic construction, when needed episodic information is missing, as we showed with our paradigm. On a functional level, Cheng et al. (2016) argued that the reason for the high chance of memory failure along with semantic construction may be that it "enables a flexible simulation of future situations" rather than to faithfully recreate one's past. Indeed, we could show that semantic construction was higher for task-irrelevant objects, which are presumably less important for future situations.

As we made half of the objects task-relevant. we made sure participants attended at least half of the objects, while the other half was encountered passively. The resulting behavioural patterns in our paradigm match the principles of the (virtual) enactment effect, which states that actively performed events are better recalled than passively observed actions (Cohen, 1989; Tuena et al., 2019; Williams et al., 2005). Accordingly, task-relevant objects were better remembered than task-irrelevant objects in our study.

In our study, we found that congruent objects had a consistently higher chance of being recalled (correctly). This effect has been found in previous studies as well (Hall & Geis, 1980; Marks et al., 1992; van Kesteren et al., 2019). van Kesteren et al. (2012) argued that if a (semantic) schema is activated by exposure to a known context, it can on one hand activate schema-driven attention processes and at the same time facilitate the organisation of new information. Our finding that the likelihood of memory recall was highest when episodic and semantic information

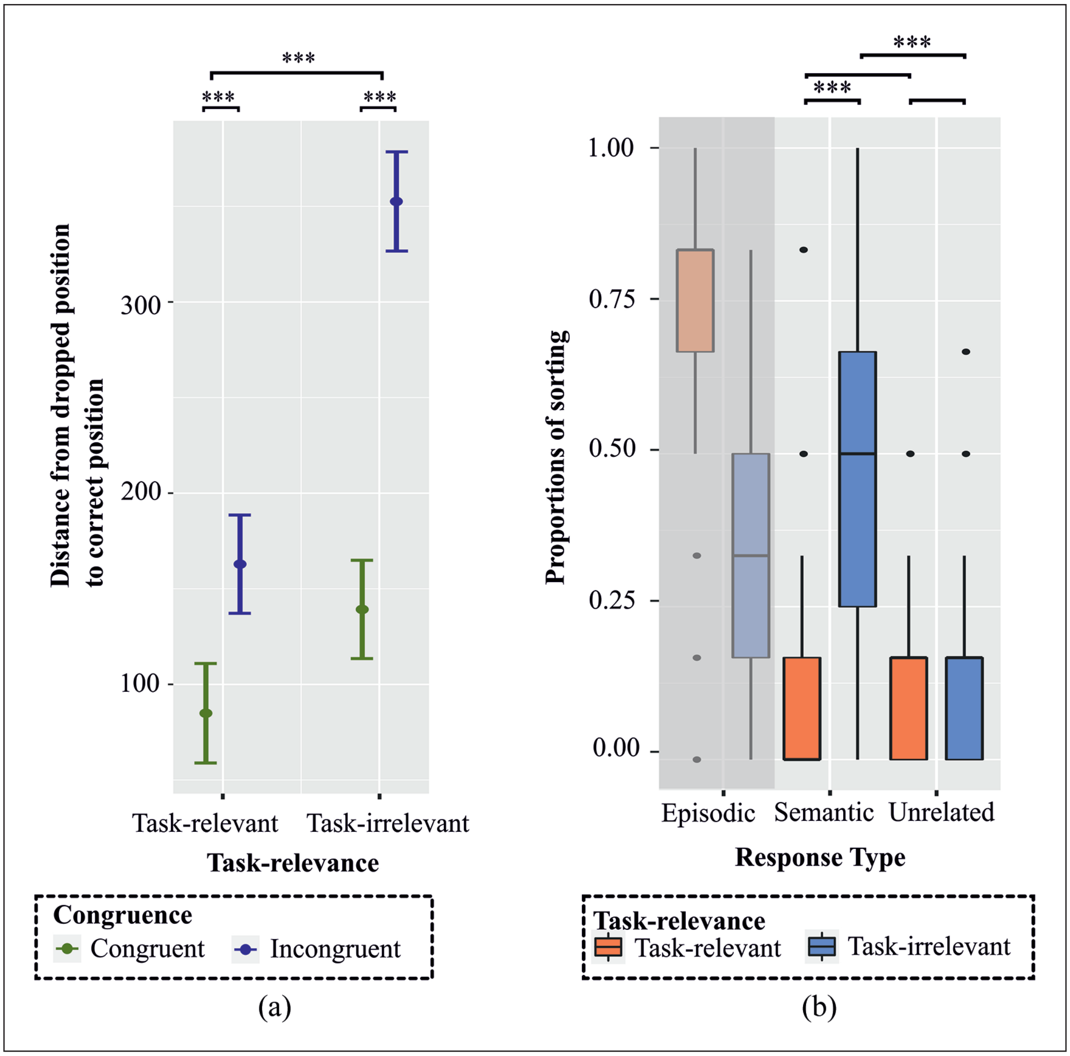

Figure 7. Findings from the replication study. (a) Model estimates and standard errors from the linear mixed model, predicting the distance from the dropped position to the correct position (drop-error) with task-relevance and congruence. (b) Results from the semantic construction analyses including only incongruent objects.

The box-plots show the median, the first and third quartiles (represented by the box), the minimum and maximum (represented by the whiskers), and outliers (data points which are smaller or larger than 1.5 times the interquartile range). In the case of episodic memory failure, task-irrelevant objects were more likely placed to the semantically fitting room, rather than to the unrelated room. This was not the case for task-relevant objects. Proportions of episodic sortings were not included in statistical analyses and are marked as shaded.

***$p < .001$.

matched (for congruent object—location combinations) is in line with this idea. Gronau et al. (2008) found that the recognition of objects already benefits from contextual associations. In a later published review, they argued that both semantic relation and the task-relevance of an object influence attentional processes and thus explain differences in perception and accordingly in memory (Gronau, 2021). Many studies regarding the congruence effect tested memory performance for verbal material, whereas in our study we investigated memory in a virtual environment with naturalistic 3D objects. Thus, we replicated the congruence effect in an incidental learning setting.

Nevertheless, using common household objects placed at odds could also have had an opposing effect on object memory: unexpected, surprising object locations could have been more noticeable and thus more easily recalled, as also reported in previous studies (Kormi-Nouri et al., 2005; Küppers & Bayen, 2014). Prior to conducting this study, we acknowledged that based on the current literature, incongruent objects could possibly be remembered worse (due to the congruence effect) or better (because of unexpectedness), compared with congruent objects. The significant interaction between congruence and task-relevance as predictors for spatial memory reflects that the congruence effect is less pronounced when participants actively attended an incongruent object, by interacting with it (see Figure 5b). Accordingly, we found that the time participants have spent searching for incongruent objects exceeded the search-time for congruent objects by almost 10 s. In our environment, half of the objects appeared incongruently, so that each encountering may have entailed a smaller prediction error. Consequently, the further absence of the effect of surprise (and the enhanced recall of congruent object locations) may be due to the high number of incongruent objects, making each additional incongruent encountering with an object less surprising. Finally, we used surprising, but logically possible object placements—all objects were transportable and thus relocatable.

One additional manipulation in our study design regarded the effect of retrieval delay. We speculated that over time, episodic memory accuracy might decrease, while the influence of semantic information on memory retrieval might increase accordingly. This assumption was based on previous findings indicating that the likelihood of episodic memory failure increases over time (Mitchell et al., 1990; Stafford et al., 1987; Sun et al., 2018). However, we did not consistently find this effect in our analyses. The retrieval delays had neither an effect on spatial memory accuracy nor on the semantic bias. One reason for this lack of effects of delay is a first limitation we identified in the repeated-measures task design we used: The first recall session included (1) a PVT, in which all encountered objects were presented multiple times and (2) a spatial recall task, where subjects were asked to sort all encountered objects to their remembered position. There is experimental evidence that a reencountering of objects leads to a re-encoding and potentially to reconsolidation (Bjork, 1975; Forcato et al., 2014; Roediger & Payne, 1982). Also, memory recall in a testing situation such as in a laboratory setting has been identified as a learning experience of itself, which is known as the "testing effect" (Toppino & Cohen, 2009). The first recall session our participants underwent hence likely influenced the second recall 1 week later, in which we cannot solely relate memory recall to the initially encoded episode. The usage of a between-subject design would be more suitable in this case. Nevertheless, the first recall session did not include any feedback on the room-sorting, spatial recall, or temporal order, so that the effects over time can cautiously be interpreted. van Kesteren et al. (2013), for example, investigated the superiority of memory for congruently presented items versus incongruently presented items over time. They found that for recognition memory, the congruence-effect was only present after a consolidation period, but not after immediate recall. Interestingly, the memory for congruent items was then more stable over time as compared with incongruent items. We could not replicate this finding in our data but believe that this might be due to the usage of a within-subject design. Future studies should investigate in more detail the development of the congruence-effect over time.

When recalling the order of the tasks we found a temporal clustering of object-pairs belonging to the same semantic category in the temporal recall compared with object-pairs belonging to different semantic categories. Importantly, this effect was only present 1 week, but not 1 day after encoding. Prior studies using word-list learning already showed semantic clustering in temporal, free recall tasks (Bousfield, 1953; Manning & Kahana, 2012). Our results extend these findings, as we could show influences of semantic information even on the explicit reconstruction of the temporal order of tasks from a realistic encoding situation in a cued recall setting.

After using a first-person perspective during encoding, we used a shift in perspective during retrieval of the spatial recall, as the flat was presented from a birds'-eye perspective. A shift in perspective from encoding to retrieval could result in impaired accuracy of memories (Marcotti & St Jacques, 2018). Nevertheless, we argue that the investigation of scenario construction is not hindered in principle by this shift in perspective. The current design could, however, be further developed by controlling for a consistent perspective during encoding and retrieval.

Our results support the notion that remembering a past event requires interplay of the episodic memory traces and semantic information, and does not rely solely on actually encoded information, as originally suggested by Tulving and Donaldson (1972). The present study thus provides a behavioural support for the SCM, which proposes that during recall, only the gist of the memory is retrieved, and missing details are constructed from semantic information. The usage of a VR has proven to be beneficial for the investigation of scenario construction to experimentally differentiate between episodic memory traces and semantic information by being able to selectively violate expectations and hence create conflict between the two memory systems. The design is very feasible, as it can be easily adapted for different settings. In addition, the data presented in this study were successfully modelled by Fayyaz et al. (2022), thus contributing to a bigger understanding of underlying processes. The application of our design combined with neuroimaging techniques will provide further insights into the neural correlates of scenario construction.

Acknowledgements

The authors gratefully acknowledge the help of Henry Soldan, Hannah Heidemeyer, and Laura Badziong during data collection and recruitment of participants. Moreover, they cordially thank Tobias Otto for technical support. In addition, they thank Anne Bierbrauer and Nora Herweg for their expertise and advice. Last but not least, they thank the anonymous reviewer for their enormously helpful feedback.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in the framework of the research unit FOR 2812 under grant nos 419039274 (O.T.W.), 419049386 (N.A.), 419037023 (R.I.S.), 419039588 (S.C.), and 419037518 (S.C.).

ORCID iD

Oliver T Wolf https://orcid.org/0000-0002-9320-2124

Data accessibility statement

The data of the present experiment and the analyses codes can be accessed using the following link: https://osf.io/jdhpr/?view_only=1df8a2b162854ea5884bb190734144fc.

Supplementary material

The supplementary material is available at qjep.sagepub.com.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: The role of integration and relatedness. *PLOS ONE*, *10*(2), Article e0115624. https://doi.org/10.1371/journal.pone.0115624

Bierbrauer, A., Fellner, M.-C., Heinen, R., Wolf, O. T., & Axmacher, N. (2021). The memory trace of a stressful episode. *Current Biology*, *31*, 5204–5213.e8. https://doi.org/10.1016/j.cub.2021.09.044

Bjork, R. A. (1975). Retrieval as a modifier. In Information processing and cognition: The Loyola symposium (pp. 123–144). Hillsdale, NJ: Erlbaum.

Blender Online Community. (2018). *Blender—A 3D modelling and rendering package* [Computer software]. Stichting Blender Foundation. http://www.blender.org

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, *49*(2), 229–240. https://doi.org/10.1080/00221309.1953.9710088

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230.

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel). GESIS—Leibniz-Institut für Sozialwissenschaften*. https://doi.org/10.6102/zis242

Brod, G., & Shing, Y. L. (2019). A boon and a bane: Comparing the effects of prior knowledge on memory across the lifespan. *Developmental Psychology*, *55*(6), 1326–1337. https://doi.org/10.1037/dev0000712

Brooks, B. M., Attree, E. A., Rose, F. D., Clifford, B. R., & Leadbetter, A. G. (1999). The specificity of memory enhancement during interaction with a virtual environment. *Memory*, *7*(1), 65–78. https://doi.org/10.1080/741943713

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180–10185. https://doi.org/10.1073/pnas.1610686113

Cheke, L. G., & Clayton, N. S. (2013). Do different tests of episodic memory produce consistent results in human adults? *Learning & Memory*, *20*(9), 491–498. https://doi.org/10.1101/lm.030502.113

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience and Biobehavioral Reviews*, *60*, 82–89. https://doi.org/10.1016/j.neubiorev.2015.11.011

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review*, *1*(1), 57–80. https://doi.org/10.1007/BF01326550

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*(3), 305–312. https://doi.org/10.2466/pr0.1959.5.3.305

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fayyaz, Z., Altamimi, A., Zoellner, C., Klein, N., Wolf, O. T., Cheng, S., Wiskott, L. (2022). A model of semantic completion in generative episodic memory. *Neural Computation, 34*(9), 1841–1870. https//:doi: 10.1162/neco_a_01520

Forcato, C., Fernandez, R. S., & Pedreira, M. E. (2014). Strengthening a consolidated memory: The key role of the reconsolidation process. *Journal of Physiology, 108*(4–6), 323–333. https://doi.org/10.1016/j.jphysparis.2014.09.001

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*(3), 324–363. https://doi.org/10.1016/0010-0285(85)90012-X

Gronau, N. (2021). To grasp the world at a glance: The role of attention in visual and semantic associative processing. *Journal of Imaging*, *7*(9), Article 191. https://doi.org/10.3390/jimaging7090191

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations.

Journal of Cognitive Neuroscience, *20*(3), 371–388. https://doi.org/10.1162/jocn.2008.20027

Hall, D. M., & Geis, M. F. (1980). Congruity and elaboration in free and cued recall. *Journal of Experimental Psychology: Human Learning & Memory*, *6*(6), 778–784. https://doi.org/10.1037/0278-7393.6.6.778

Hardiess, G., Mallot, H. A., & Meilinger, T. (2015). Virtual reality and spatial cognition. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 133–137). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.43098-9

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). "Active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 383–390. https://doi.org/10.3758/BF03195466

Jebara, N., Orriols, E., Zaoui, M., Berthoz, A., & Piolino, P. (2014). Effects of enactment in episodic memory: A pilot virtual reality study with young and elderly adults. *Frontiers in Aging Neuroscience, 6*, Article 338. https://doi.org/10.3389/fnagi.2014.00338

Kormi-Nouri, R., Nilsson, L.-G., & Ohta, N. (2005). The novelty effect: Support for the novelty-encoding hypothesis. *Scandinavian Journal of Psychology*, *46*(2), 133–143.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language*, *37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology*, *67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lenth, R. V. (2020). *emmeans: Estimated marginal means, aka least-squares means* [Computer software]. https://CRAN.R-project.org/package=emmeans

Manning, J. R., & Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, *20*(5), 511–517. https://doi.org/10.1080/09658211.2012.683010

Marcotti, P., & St Jacques, P. L. (2018). Shifting visual perspective during memory retrieval reduces the accuracy of subsequent memories. *Memory*, *26*(3), 330–341. https://doi.org/10.1080/09658211.2017.1329441

Marks, W., McFalls, E. L., & Hopkinson, P. L. (1992). Encoding pictures in scene contexts: Does task demand influence effects of encoding congruity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 192–198.

MATLAB. (2020). *version 7.10.0 (R2020a)* [Computer software]. The MathWorks Inc.

Mitchell, D. B., Brown, A. S., & Murphy, D. R. (1990). Dissociations between procedural and episodic memory: Effects of time and aging. *Psychology and Aging*, *5*(2), 264–276. https://doi.org/10.1037/0882-7974.5.2.264

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Pacheco, E. D., Zucca, R., Arsiwalla, X., Principe, A., Zhang, H., Rocamora, R., Axmacher, N., & Verschure, P. F. (2021). Volitional learning promotes theta phase coding in the human hippocampus. *Proceedings of the National Academy of Sciences*, *118*(10), Article e2021238118.

Palombo, D. J., Williams, L. J., Abdi, H., & Levine, B. (2013). The survey of autobiographical memory (SAM): A novel measure of trait mnemonics in everyday life. *Cortex*, *49*(6), 1526–1540. https://doi.org/10.1016/j.cortex.2012.08.023

R Core Team. (2019). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org/

Raybaut, P. (2009). *Spyder* [Computer software]. pythonhosted.org

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research* [Computer software]. https://CRAN.R-project.org/package=psych

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Roediger, H. L., & Payne, D. G. (1982). Hypermnesia: The role of repeated testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(1), 66–72. https://doi.org/10.1037/0278-7393.8.1.66

Rose, J., Otto, T., & Dittrich, L. (2008). The Biopsychology-Toolbox: A free, open-source MATLAB-toolbox for the control of behavioral experiments. *Journal of Neuroscience Methods*, *175*(1), 104–107. https://doi.org/10.1016/j.jneumeth.2008.08.006

RStudio Team. (2019). *RStudio: Integrated development environment for R* [Computer software]. http://www.rstudio.com/

Sauzéon, H., Arvind Pala, P., Larrue, F., Wallet, G., Déjos, M., Zheng, X., Guitton, P., & N'Kaoua, B. (2011). The use of virtual reality for episodic memory assessment: Effects of active navigation. *Experimental Psychology*, *59*(2), 99–108. https://doi.org/10.1027/1618-3169/a000131

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Sipe, S. J., & Pathman, T. (2020). Memory at play: Examining relations between episodic and semantic memory in a children's museum. *Child Development*, *92*, e270–e284.

Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*(1), 203–218. https://doi.org/10.5334/irsp.90

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall, mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research*, *14*(2), 203–229. https://doi.org/10.1111/j.1468-2958.1987.tb00127.x

Staresina, B. P., Gray, J. C., & Davachi, L. (2009). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*(5), 1198–1207. https://doi.org/10.1093/cercor/bhn165

Sun, Q., Gu, S., & Yang, J. (2018). Context and time matter: Effects of emotion and motivation on episodic memory overtime. *Neural Plasticity, 2018*, Article 7051925. https://doi.org/10.1155/2018/7051925

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252–257. https://doi.org/10.1027/1618-3169.56.4.252

Tuena, C., Serino, S., Dutriaux, L., Riva, G., & Piolino, P. (2019). Virtual enactment effect on memory in young and aged populations: A systematic review. *Journal of Clinical Medicine*, *8*(5), Article 620. https://doi.org/10.3390/jcm8050620

Tulving, E. (2001). Episodic memory and common sense: How far apart? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *356*(1413), 1505–1515. https://doi.org/10.1098/rstb.2001.0937

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Tulving, E., & Donaldson, W. (1972). *Organization of memory*. Academic Press.

van Kesteren, M. T. R., Rignanese, P., Gianferrara, P. G., Krabbendam, L., & Meeter, M. (2019). *Integrating memories: Congruency and reactivation aid memory integration through reinstatement of prior knowledge*. https://doi.org/10.1101/716076

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2013). Consolidation differentially modulates schema effects on memory for items and associations. *PLOS ONE*, *8*(2), Article e56155. https://doi.org/10.1371/journal.pone.0056155

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual* [Computer software]. CreateSpace.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, *67*(5), 816–827. https://doi.org/10.3758/bf03193535