Article:

Andermane, Nora, Moccia, Arianna, Zhai, Chong et al. (2 more authors) (2025) The holistic forgetting of events and the (sometimes) fragmented forgetting of objects. Cognition. 106017. ISSN 0010-0277

https://doi.org/10.31234/osf.io/tb67g

Cognition 255 (2025) 106017

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

The holistic forgetting of events and the (sometimes) fragmented forgetting of objects

Nora Andermane [a,b], Arianna Moccia [a], Chong Zhai [a], Lisa M. Henderson [a,c], Aidan J. Horner [a,c,*]

[a] *Department of Psychology, University of York, UK*
[b] *School of Psychology, University of Sussex, UK*
[c] *York Biomedical Research Institute, University of York, UK*

ARTICLE INFO

Keywords:
Episodic memory
Event memory
Object memory
Retrieval dependency
Forgetting

ABSTRACT

Episodic events are typically retrieved and forgotten holistically. If you recall one element (e.g., a person), you are more likely to recall other elements from the same event (e.g., the location), a pattern that is retained over time in the presence of forgetting. In contrast, representations of individual items, such as objects, may be less coherently bound, such that object features are forgotten at different rates and retrieval dependency decreases across delay. To test the theoretical prediction that forgetting qualitatively differs across levels in a representational hierarchy, we investigated the potential dissociation between event and item memory across five experiments. Participants encoded three-element events comprising images of famous people, locations, and objects. We measured retrieval accuracy and the dependency between the retrieval of event associations and object features, immediately after encoding and after various delays (5 h to 3 days). Across experiments, retrieval accuracy decreased for both events and objects over time, revealing forgetting. Retrieval dependency for event elements (i.e., people, locations, and objects) did not change over time, suggesting the holistic forgetting of events. Retrieval dependency for object features (i.e., state and colour) was more variable. Depending on encoding and delay conditions across the experiments, we observed both fragmentation and holistic forgetting of object features. Our results suggest that event representations remain coherent over time, whereas object representations can, but do not always, fragment. This provides support for our representational hierarchy framework of forgetting, however there are (still to be determined) boundary conditions in relation to the fragmentation of object representations.

1. Introduction

Real-world events are complex in nature and are composed of varied multisensory information about where we were, who we were with, and what we were doing. Although we are often able to remember such events in vivid detail (Tulving, 1983), recollecting the constituent parts of the event precisely and accurately (Wagenaar, 1986), we also forget over time (Ebbinghaus, 1885; Murre & Dros, 2015). Sometimes we might forget the entire event, and other times we might remember most of the event elements (e.g., the location and person), but forget specific perceptual details. For example, we might remember a friend giving us a present in a coffee shop but forget the colour of the wrapping paper. Forgetting can therefore vary across events and can occur at multiple levels – from perceptual details to higher-level event elements (Andermane, Joensen, & Horner, 2021; Fisher & Radvansky, 2019; Sacripante, Logie, Baddeley, & Della Sala, 2022; Sadeh, Ozubko, Winocur, & Moscovitch, 2014).

Research on forgetting has typically focused on the rate and form of forgetting (Berens, Richards, & Horner, 2020; Ebbinghaus, 1885; Rubin & Wenzel, 1996) as well as whether forgetting occurs via interference or decay (Jenkins & Dallenbach, 1924; McGeoch, 1932; Underwood, 1957). More recently, several studies have asked whether forgetting is 'holistic' in nature – when we forget one detail of an event do we also forget other details of the same event (Brady, Konkle, Alvarez, & Oliva, 2013; Joensen, Gaskell, & Horner, 2020)? For example, if we experience an event consisting of a location, person, and object, do we forget holistically such that we forget all three elements together, or does the event representation fragment over time such that we forget some elements but not others? Similarly, if we encounter an object with multiple features (e.g., shape, colour, state etc.), do we forget all these features

* Corresponding author at: Department of Psychology, University of York, Heslington, York YO10 5DD, UK.
E-mail address: aidan.horner@york.ac.uk (A.J. Horner).

https://doi.org/10.1016/j.cognition.2024.106017
Received 24 May 2024; Received in revised form 19 November 2024; Accepted 20 November 2024
Available online 29 November 2024

together, or do we forget some features but not others (e.g., we remember the state but not the colour)? Regardless of whether we are discussing events (and the elements that make up an event) or objects (and the features that make up an object), the central question is whether mnemonic representations are forgotten holistically, in an "all-or-none" manner, or instead fragment over time.

In relation to events, and the elements that make up an event, we have recently provided evidence that forgetting is holistic in nature. If we forget one element of an event (e.g., the location), we tend to also forget the other elements (e.g., the person and object) (Joensen et al., 2020). Participants learnt individual elements (i.e., locations, objects, and people) as overlapping word pair associations (*kitchen-hammer*, *hammer-Barack Obama*, *Barack Obama-kitchen*) and were asked to imagine the associations vividly, such that across three encoding trials the three elements formed a coherent episodic 'event'. Immediately after encoding, and after various delays, participants' memory of the event associations was tested. We used a measure of retrieval dependency – whether one element is more likely to be retrieved given another element from the same event is also retrieved – to assess whether forgetting is holistic. We have previously used retrieval dependency to demonstrate that event retrieval is holistic and that it is underpinned by a hippocampal pattern completion process (Grande et al., 2019; Horner, Bisby, Bush, Lin, & Burgess, 2015; Horner & Burgess, 2013, 2014). To assess forgetting, we measured retrieval dependency immediately after encoding and after a delay, where forgetting has occurred. If retrieval accuracy decreases over time (i.e., forgetting takes place) but retrieval dependency remains consistent, this would suggest that event elements are forgotten holistically. However, if retrieval dependency decreases over time (alongside decreases in accuracy), this would suggest that event elements are forgotten in a more fragmented manner. Across four experiments, we found no evidence for decreases in retrieval dependency, despite significant forgetting (Joensen et al., 2020). Thus, we provided evidence for holistic forgetting of event elements.

In support of the proposal that event elements are both encoded and forgotten in a holistic manner, Cooper and Ritchey (2019) assessed memory for spatial context and object colour using a 'precision' memory paradigm (Cooper & Ritchey, 2019). They found retrieval dependency for spatial context and object colour in relation to 'gist' accuracy (i.e., whether participants remembered both elements with a certain degree of accuracy), however the precision with which they remembered each element were unrelated. This suggests the binding of event elements might occur at a more general 'gist' or 'semantic' level (e.g., a mug in an office) and that the features of each element are represented more independently (e.g., a *green* mug in a *small* office). Similar questions have been asked in the source memory literature, where retrieval dependency is assessed for different 'sources' of a visually presented word – for example the location and font of the word – with evidence both for and against holistic binding (Meiser & Bröder, 2002; Starns & Hicks, 2005, 2008). Importantly, the Cooper and Ritchey (2019) and source memory studies did not look at forgetting via changes in retrieval dependency over time (as in Joensen et al., 2020 and the present experiments).

While event elements are forgotten in a holistic (all-or-none) manner, this may not be the case for the features that make up an object. For example, object features like an object's exemplar, state, and colour have been found to fragment as a function of forgetting (Brady et al., 2013). First, the state and colour of an object were shown to be forgotten at different rates. Second, retrieval dependency was shown to decrease after a delay; the successful retrieval of one object feature (e.g., exemplar) no longer predicted the retrieval of another feature (e.g., state). Based on these findings Brady and colleagues concluded that object representations fragment over time – the different features of an object were not bound together at encoding and were therefore forgotten independently of each other. This independent encoding of object features is further supported by Utochkin and Brady (Utochkin & Brady, 2020). If participants learn two similar objects (e.g., a mug) which differ in both exemplar and state (e.g., full or empty), they can successfully remember exemplars and states, but struggle to remember which exemplar was in which state. Further, following the encoding of a specific object exemplar in a specific state, showing the exemplar in a different incongruent state at test did not significantly decrease retrieval accuracy. This was not the case when manipulating hue and brightness, two features that are integral to the perception of colour and therefore are more likely to be bound. Although not focused on forgetting, these results strengthen the proposal that certain object features are encoded and forgotten independently.

However, there have also been contrasting findings suggesting that object features may be forgotten holistically. Using a different statistical method to assess retrieval dependency from Brady et al. (2013), Balaban and colleagues (Balaban, Assaf, Arad Meir, & Luria, 2020) provided evidence for holistic forgetting across several different features: exemplar, state, material, and orientation. They also varied the nature of the encoding task and the duration of stimulus presentation time at encoding. Across five experiments they found no evidence for fragmented forgetting. Critically, they always tested memory for exemplars relative to either state, material, or orientation and their multinominal modelling approach provided evidence for a hierarchical representation. Remembering the object exemplar was predictive in relation to remembering state, material, or orientation, but remembering one of these latter features wasn't necessarily predictive of remembering exemplar. Thus, it is possible that the lack of fragmentation in Balaban et al. (2020) is a function of testing object exemplar. Note, however, that this can't fully explain why Brady et al. (2013) did find evidence of fragmentation given they tested object exemplar and state.

One proposal is that objects and their features can be represented both holistically and/or independently (Kuhbandner, 2020), dependent on the encoding and task conditions. For example, Spachtholz and Kuhbandner (2017) provided evidence that the emotional affect of the participant at encoding modulated the extent to which object features were holistically represented. Specifically, positive affect increased retrieval dependency, suggesting more holistic representations, relative to negative affect. Further support for this proposal comes from Li et al. (2022) who showed that object features might be bound at a 'lower resolution' (e.g., green) but represented with higher resolution (e.g., this specific green) in an independent manner (Li, Fukuda, & Barense, 2022). Thus, object representations might be hierarchical in manner (Brady, Konkle, & Alvarez, 2011), with precise representations of features represented independently and less precise representations of features being bound together further up the representational hierarchy.

Although evidence for fragmentation of object features is mixed, it is possible that fragmentation does occur under certain conditions and evidence for object features being represented more holistically (Ceraso, Kourtzi, & Ray, 1998; Duncan, 1984; Vogel, Woodman, & Luck, 2001; Wilton, 1989) and more independently (Oberauer & Eichenberger, 2013; Olson & Jiang, 2002; Reinitz, Lammers, & Cochran, 1992) in both short-term and long-term memory exist in the literature. It is therefore likely that, under specific conditions, object features are represented and forgotten independently, in contrast to the more holistic forgetting of event elements. To explain this possible dissociation between object-based and event-based representations, we recently proposed a model where forgetting qualitatively differs across different levels in a representational hierarchy (Andermane et al., 2021). Specifically, we proposed that objects and their features are represented across the ventral visual stream, with individual features represented independently in distinct regions (Grill-Spector, Kourtzi, & Kanwisher, 2001; Martin, Douglas, Newsome, Man, & Barense, 2018) that converge on the perirhinal cortex (Barense, Gaffan, & Graham, 2007; Bussey & Saksida, 2007). As these object features are represented independently, we reasoned they would be more likely to show fragmented forgetting. Specifically, feature-specific interference (Sadeh et al., 2014; Sadeh, Ozubko, Winocur, & Moscovitch, 2016) might induce forgetting for one object feature without impacting memory for the other features of the object. We also proposed that event elements were bound more holistically in the (posterior) hippocampus (Cohen & Eichenbaum, 1993; Horner & Doeller, 2017; Marr, 1971; Mayes, Montaldi, & Migo, 2007; Scoville & Milner, 1957; Teyler & DiScenna, 1986), resulting in a more holistic form of forgetting. Here, due to more pattern separated representations in the hippocampus (Bakker, Kirwan, Miller, & Stark, 2008; Berron et al., 2016; LaRocque et al., 2013; Motley & Kirwan, 2012; Neunuebel & Knierim, 2014; Yassa & Stark, 2011), holistic forgetting is more likely a consequence of decay (Hardt, Nader, & Nadel, 2013; Sadeh et al., 2014).

In the present series of experiments, we tested the hypothesis that events and their constituent elements are forgotten holistically, whereas objects and their features can be forgotten in a more fragmented manner. To our knowledge, one previous study has assessed this possible distinction using novels that contained both event elements and features of objects, finding no evidence for decreases in dependency for either level of representation over time (Parra, Antes, & Radvansky, 2024). Here we used visual stimuli (images of locations, people, and objects) to (1) test the boundary conditions for the lack of fragmentation seen previously for events (Joensen et al., 2020; Parra et al., 2024) and (2) make our object-level results more comparable to the object memory literature (Balaban et al., 2020; Brady et al., 2013). While the distinction between holistic and fragmented forgetting is binary – no decrease in dependency is evidence for holistic forgetting and any decrease is evidence for fragmented forgetting – fragmentation itself can vary along a continuum. You can have more or less fragmentation, leading to larger vs smaller decreases in dependency over time. For simplicity, we refer to holistic vs fragmented forgetting throughout the manuscript, focussing on the presence vs absence of evidence for fragmentation, rather than the amount of fragmentation.

Although evidence exists for holistic forgetting of events (Joensen et al., 2020) and fragmented forgetting of objects (Brady et al., 2013), this potential dissociation has not been tested in a single experimental framework using the same stimuli. Critically, whereas Joensen et al. (2020) used written word stimuli, Brady et al. (2013) used images of objects. Thus, any potential dissociation can currently be explained by stimulus differences. We therefore created a novel experimental approach where participants encoded 'events' with three elements – a location, famous person, and object – where each element was a visual image. Further, each object within an event had two distinct features – colour and state. This allowed us to assess memory, and retrieval dependency, for event elements (e.g., if we cue with location can you remember the object?) and object features (e.g., can you remember which state and colour this object was presented in?) while keeping encoding conditions consistent. We only tested memory for object features, and not location or person features, to be more comparable with the object memory literature and because we were able to use the same stimulus set as in Brady et al. (2013) where the colour and state of objects were systematically manipulated.

We tested memory across several delays to assess retrieval dependency for event elements and object features as a function of forgetting. We aimed to replicate the findings of Joensen et al. (2020), showing that event elements were forgotten holistically (even when using visual rather than verbal stimuli), and to further investigate the possible fragmentary nature of forgetting for objects. We did not test the prediction of our hierarchical framework that object features are forgotten via interference whereas event elements are forgotten via decay, however we return to this topic in the general discussion. Although the manner of testing event elements and object features differs somewhat in our experiments, to be as similar as possible to the manner in which Joensen et al. (2020) and Brady et al. (2013) tested events and objects respectively, the encoding conditions as are similar as possible (identical in Experiment 3) between conditions, and thus our experiments represent the first systematic attempt to directly test for differences between events and objects in relation to holistic vs fragmented forgetting.

Across five experiments (Experiments 1–4 were preregistered: Experiment 1 https://osf.io/vhtc3, Experiment 2 https://osf.io/gxdtr, Experiment 3 https://osf.io/kjq4u, Experiment 4 https://osf.io/32zwm, Experiment 5 was not preregistered), we provide evidence for holistic forgetting of events, and evidence for both fragmentation and holistic forgetting of objects. Experiment 1 assessed forgetting of event elements, providing evidence for holistic forgetting. Experiment 2 assessed forgetting of object features, providing evidence for fragmentation. Experiment 3 assessed both event elements and object features in a within-subject design, providing evidence of holistic forgetting for both events and objects. Experiment 4 provided further evidence for holistic forgetting of objects when objects were tested across two delay periods. In Experiment 5, we only presented object stimuli, and varied encoding duration, providing evidence for fragmentation when encoding time was reduced. Although we are not able to identify the precise boundary conditions under which object features fragment, our results reconcile previous discrepancies in the literature, suggesting objects can be (though are not always) forgotten in a fragmented manner. More broadly, the holistic forgetting of events and the (sometimes) fragmented forgetting of objects provides support for our hierarchical model of forgetting (Andermane et al., 2021).

2. Experiment 1

Experiment 1 assessed whether the retrieval of event elements (i.e., locations, people, and objects) remains holistic in the presence of forgetting. We assessed retrieval accuracy and dependency immediately after encoding and after a 3-day delay. We predicted that accuracy would decrease across delay, consistent with the presence of forgetting, whereas retrieval dependency would not decrease, consistent with the presence of holistic forgetting. This pattern of results would replicate those of (Joensen et al., 2020) but for visual images (images of locations, people, and objects) rather than verbal stimuli (written words).

2.1. Methods

2.1.1. Participants

Fifty-six participants were recruited for this two-part online experiment via Prolific (www.prolific.co). All participants across Experiments 1–5 provided informed consent and all experiments were approved by the University of York Psychology Department Ethics Committee. To be eligible for recruitment, participants had to be native English speakers (aged 18 to 40) who have lived in the UK for 5 or more years, with normal or corrected-to-normal vision and no colour-blindness. Participants were further excluded if memory performance was <30 % or >95 % in the immediate testing condition. Of the participants who started the experiment, 39 participants (Mean age = 30, SD = 6.03; 7 male, 32 female) completed both sessions, met the pre-registered inclusion criteria and were retained in the analyses. The other 17 participants (30 %) were excluded due to one of the following reasons: not returning for the final session (4), having an immediate retrieval accuracy at or above 95 % (8), or having an immediate retrieval accuracy at or below 30 % (5). Note, this high exclusion rate relates primarily to retrieval accuracy (floor and ceiling effects).

Although accuracy is an important dependent variable for assessing the presence of forgetting, it is not the central dependent variable. Retrieval dependency is the critical theoretically relevant variable that we use to draw conclusions related to holistic vs fragmented forgetting. To accurately estimate retrieval dependency, we need sufficient variation across trials, and thus floor and ceiling effects affect our ability to achieve this. If participants are not at floor or ceiling, retrieval dependency is independent of retrieval accuracy and thus excluding based on accuracy is highly unlikely to bias our effects in relation to retrieval dependency.

We conducted a power analysis based on six published studies (Bisby, Horner, Bush, & Burgess, 2018; Horner et al., 2015; Horner &

Burgess, 2013, 2014; Joensen et al., 2020; Ngo, Horner, Newcombe, & Olson, 2019) that all presented complex episodic events consisting of three elements (object, person, location) at encoding. This analysis included 22 experiments where dependency (between remembering the identity of event elements) was assessed in the data relative to the model of independent retrieval. The effect sizes for this comparison, as well as the corresponding sample sizes, were collated. The power analysis was conducted separately for experiments where dependency was assessed immediately after encoding (16 experiments) versus after a week delay (3 experiments) and the effect sizes were weighted by the sample of each experiment. Overall, the power analysis indicated that for 95 % power to detect a significant retrieval dependency ($\alpha = 0.05$) with an effect size of Cohen's $d = 0.95$, a sample of $N = 16$ was required. This estimate was the same for both immediate and week delayed tests.

We also conducted a power analysis based on the study by Brady et al. (2013) who investigated forgetting of object features, as we intended to test the same effects in further experiments. We first estimated the required sample to detect retrieval dependency between remembering colour given state and state given colour immediately after encoding. We used the smallest effect immediately after encoding, i.e., dependency for remembering state given exemplar ($d = 1.2$) to inform this analysis. Detecting dependency immediately after encoding with 95 % power ($\alpha = 0.05$) would require a sample of $N = 12$. We then calculated the required sample to detect a change in dependency after a delay with a power of 95 % (using the effect size for the change in dependency of remembering the exemplar given state after 3 days ($d = 0.75$) ($\alpha = 0.05$). The recommended sample size for detecting this effect was $N = 26$. Since the smallest effect size of interest in Brady et al. (2013) was based on a sample of only 13 people, it may be that the true effect in the population is smaller, with lower 95 % CI being $d = 0.2$. In the case of it being smaller (e.g., $d = 0.5$), we would still be able to detect this effect with 80 % power in a sample of 34 participants. We therefore aimed to acquire approximately $N = 30$ (usable datasets).

We over-recruited in the present study ($N = 39$ usable datasets). This was due to the fact that we were testing online with a delayed final test and it was difficult to estimate how many participants to recruit to ensure the correct number of usable datasets following exclusions. For the full pre-registration of this experiment, see: https://osf.io/vhtc3. Note, we preregistered and piloted ($N = 11$) a similar version of this experiment (https://osf.io/cvqre) prior to this experiment. The longer encoding times of the first experiment resulted in performance being at ceiling. We therefore stopped data collection on this first experiment, changed encoding presentation time to decrease memory performance, and preregistered a new study which we carried out to completion.

2.1.2. Materials

The memory task was designed using the experiment builder of the online platform Gorilla (https://app.gorilla.sc/). Each event consisted of three elements: an image of an object, a location, and a famous person. For the purposes of creating the events, sixty images of objects were obtained from the database of real-world objects used by Brady et al. (2013), sixty images of scenes were obtained from a database of scenes (Konkle, Brady, Alvarez, & Oliva, 2010) and additional google image searches, and sixty images of famous people were obtained via Google image searches (see http://bradylab.ucsd.edu/stimuli.html for the full database of objects and scenes).

All object images were originally without background and the background was removed from the images of the celebrities (via the webpage www.remove.bg). Each of the sixty object identities were manipulated in terms of colour and state, creating four images of each object. The manipulation of state differed across objects, depending on the type of object being tested. Examples included a coffee cup that is full or empty, a toolbox with the lid open or closed, an analogue clock showing different times, an umbrella that is open or closed, and a chess board at different orientations. Two colours and two states per object were crossed resulting in an object with colour A and state A, colour A and state B, colour B and state A, and colour B and state B (the manipulation of colour and state was used to test object memory in Experiment 2, but was included here to ensure the experiments were as similar as possible). One of these four object images was chosen randomly for each of the 60 objects, which was then assigned to one of the events to be presented at encoding. The colours of the objects used in the study were carefully selected from the database so that the proportion of each colour was balanced in relation to the others. All images were cropped or re-sized to be 256 × 256 in size.

2.1.3. Design

Experiment 1 had a within-subjects design with a factor of Delay (two levels: Test phase T1 and T2). T1 was conducted immediately after encoding (following a brief break where the retrieval task was explained) and T2 was conducted after 3 days. The dependent measures of interest for the pre-registered analyses were retrieval accuracy (proportion) and retrieval dependency (see below) of event elements.

2.1.4. Procedure

The experiment consisted of an immediate session comprising a study phase and immediate test phase (T1), and a session 3 days later comprising a delated test phase (T2). Participants were recruited online via Prolific and provided informed consent before commencing with the study. A demographic questionnaire, detailed on-screen instructions, and one study phase practice trial were presented before participants could proceed to the study phase.

For an illustration of the basic trial structure and the stimuli used for event memory tasks, see Fig. 1. The same encoding structure (with variations in encoding time) was used across Experiments 1–4. The same test structure was used to assess event memory in Experiments 1 and 2. In the study phase, participants learnt 60 novel events consisting of three elements (an object, a location, and a famous person) by viewing images of the elements on the screen and then imagining these interacting in a meaningful way to form an episodic event. On each study trial,

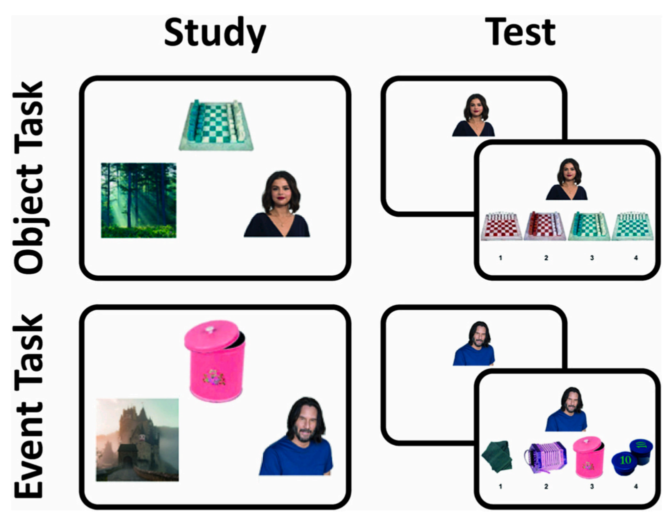

Fig. 1. Schematic of study and test trials across experiments 1–4.
At Study, for both the object and event task, triplets of locations, famous people, and objects were presented. Participants were required to pay attention to each triplet and imagine them interacting in a meaningful way. At Test, for the object task, we presented either the person or location as a cue, and participants had to select the associated object among 3 foils that manipulated both the colour and state of the object. For the event task, we presented either the location, person, or object as a cue, and participants had to select the correct associated item (location, person, or object) among 3 foils of other items of the same stimulus type that were associated with other triplets. See Methods of Experiments 1–4 for experiment specific details. Experiment 5 only presented objects at Study and only tested individual objects (with no associated location or person cue) at Test (see Experiment 5 Methods).

participants first saw a fixation point on the screen for 500 ms, followed by three images (e.g., telephone, beach, Beyoncé) for 2.5 s. The three event elements were always presented in the same locations to ensure consistency across encoding trials and participants, with the object in the top and middle of the screen, the location in the bottom left, the person in the bottom right (Fig. 1). This was followed by a blank screen with the instruction "Imagine…" for 4 s. During this final 4 s period participants were instructed to "imagine the object, the location, and the person you just saw interacting in some meaningful way". Critically, they were encouraged to "visualise an event involving the three *images* you see" to encourage visualisation of all three elements with the visual properties presented on the screen (i.e., the colour and state of the object) without drawing overt attention to the features of the objects. These instructions were designed to encourage attention to be paid to both the event elements and the visual properties of each element without focusing their attention specifically on the objects (and their features).

There was a blank inter-trial interval of 500 ms before the onset of the next trial. The study phase consisted of 60 trials, separated into four blocks. Participants were given an opportunity to take a break in between the blocks. The events were randomised, such that there were 5 study lists with different elements making up each event; the participants were allocated evenly to each study list. The events presented within each of the four blocks in the study phase was fixed for each study list but the order of the events within a block and the blocks were presented randomly intermixed for each participant.

In the test phase of the episodic memory task, participants' memory for the learnt event associations was tested by presenting one element from the event as a cue (e.g., beach) and asking them to select the associated target element (e.g., Beyoncé) among three foils from the same category (i.e., other famous people from the studied events). The test phase started with on-screen instructions and one practice trial. Participants were tested on each of the six possible element associations of each event (i.e., cue location and retrieve object, cue object and retrieve location, cue person and retrieve object etc.). Half of the events were tested at T1 (180 trials) and half were tested at T2 (180 trials). Within each session, the trials of the test phase were completed in six blocks of 30 trials, with an opportunity to take a break in between the blocks. The six blocks were necessary because each event was tested for each of the six possible cue-target associations. There were 5 versions for the test phase, corresponding to the 5 study lists where each list had 60 events made up of a new (random) combination of object, location, person elements. Within each version of the test phase, the event associations tested in each block were fixed but the order of presentation of these within a block and the block order was randomised.

On each test trial, participants saw a fixation point on the screen for 500 ms, then the cue element image appeared alone for 2 s, followed by the four response option images appearing below the cue. One of the response options was the target involved in the same event as the cue, and other three were foils of the same category belonging to other events within that same testing session (i.e., immediate or delayed). The order of the available response options on the screen from left to right was random. The options remained on the screen for 4 s or until participants made a response by pressing one of the four keys (1–4). After a 500 ms blank interval participants saw a prompt "How confident are you in your choice?" on the screen, and again were asked to press the keys 1–4 (1 - not confident, 2 - slightly confident, 3 - moderately confident, 4 - very confident) to rate their confidence. The confidence rating prompt remained on the screen for 4 s or until response. There was a blank inter-trial interval of 500 ms before the onset of the next trial.

After T2, participants were asked to report whether they had prior knowledge of the famous people presented in the memory task via a celebrity check task. An image of each famous person from the experiment was presented on the screen for 4 s or until response and participants pressed one of two keys to indicate if they know the celebrity or not. Participants also completed a Vividness of Visual Imagery (VVIQ) questionnaire (Marks, 1973). Analyses related to the VVIQ measures will be reported in a separate publication. Participants were asked to leave any feedback about the experiment in a comments box after completing each of the sessions (i.e., T1 and T2) and were reimbursed for their time at a rate of £8–9 per hour via Prolific.

If participants did not meet the inclusion criteria based on their performance in the episodic memory task at T1, they were still reimbursed for their time but were not invited back for T2. The exclusion criteria were as follows. For T1, participants were excluded from the study if they had accuracy of <30 % or > 95 %. We did not exclude any participants based on T2 performance.

2.1.5. Analyses

2.1.5.1. Retrieval accuracy and confidence. Mean retrieval accuracy (proportion) and mean retrieval confidence were calculated for the test phase of the episodic memory task at T1 and T2. As detailed in the preregistration, timeouts on memory trials were converted to incorrect responses for the purposes of calculating retrieval accuracy and dependency. At T1, 3.6 % of the memory trials (SD = 3.6 %) were timeouts, whereas at T2 4.2 % of the trials (SD = 5.2 %) were timeouts.

We also calculated what proportion of the celebrities the participants had knowledge of. On average, participants reported knowing 83 % of the famous people that were presented in the episodic events (SD = 20 %). This gave us confidence that participants had sufficient semantic knowledge of the people we presented to be able to recognise them and imagine episodic events involving them at encoding.

2.1.5.2. Retrieval dependency. Retrieval dependency in the episodic memory task was calculated as in Joensen et al. (2020). For each participant, we created six 2 × 2 contingency tables for the observed data and independent model (see Table 1). The four cells for each table for the observed data reflected the count of retrieval trials where two elements (e.g., person and object) were retrieved when cued by the other element (e.g., location; A_BA_C) or where one element (e.g., location) was retrieved when cued by the two other elements (on separate retrieval trials; e.g., person and object; B_AC_A), where the two separate retrieval trials of interest can either both be correct, both be incorrect, or one correct and one incorrect in either direction. Once constructed, we calculated the proportion of events where either both retrieval trials were correct or incorrect – called the proportion of joint retrieval (and non-retrieval).

Given this joint retrieval measure scales with accuracy, we constructed an independent model that predicts the level of joint retrieval for a given participant if retrieval trials within an event were not statistically related (Table 1). We calculated the probability, across all events, of (e.g.,) retrieving B when cued by A (P_{AB}) and retrieving C when cued by A (P_{AC}). For the A_BA_C contingency table, the probability of

Table 1
Contingency table for the independent model for correct and incorrect retrieval, over N events ($i = 1$ to N), for elements B and C when cued by A.

	Retrieval of Element B	
Retrieval of Element C	Correct (P_{AB})	Incorrect ($1 - P_{AB}$)
Correct (P_{AC})	$\sum_{i=1}^{N} = P_{AB_i} P_{AC_i}$	$\sum_{i=1}^{N} = P_{AC_i} (1 - P_{AB_i})$
Incorrect ($1 - P_{AC}$)	$\sum_{i=1}^{N} = P_{AB_i} (1 - P_{AC_i})$	$\sum_{i=1}^{N} = (1 - P_{AB_i})(1 - P_{AC_i})$

For a given participant, the proportion of correct retrievals of, for instance, element B when cued by A is denoted by P_{AB} (i.e., the mean performance for B when cued by A across all events). For the independent model, when cued by A, the probability of (1) correctly retrieving B and C (across all events) is equal to $P_{AB}\,P_{AC}$;(2) correctly retrieving B but not C is equal to $P_{AB}\,(1 - P_{AC})$;(3) correctly retrieving C but not B is equal to $P_{AC}\,(1 - P_{AB})$;and (4) incorrectly retrieving both B and C is equal to $(1 - P_{AB})\,(1 - P_{AC})$.

retrieving both B and C correctly, when cued by A, if the two are statistically independent, is $P_{AB}*P_{AC}$ (see Table 1 for the calculation for the remaining 3 cells in the contingency table). Once constructed, we calculated the proportion of joint retrieval for the independent model.

To calculate a measure of retrieval dependency for each participant we took the mean proportion of joint retrieval in the data across the six contingency tables and the same mean for the independent model and calculated the difference between these two means (observed joint retrieval – independent model joint retrieval). This resulted in a single measure of retrieval dependency at T1 and T2 per participant.

A dependency measure of zero reflects no evidence for retrieval dependency and scores greater than zero reflect evidence for retrieval dependency. This measure has no clear upper bound (e.g., it does not scale between 0 and 1), and therefore the magnitude of the measure cannot be used to infer whether dependency is maximal or not. In terms of theoretical inference, any evidence for dependency greater than zero at a single timepoint we take as evidence for some degree of holistic retrieval. Critically, we use the difference between dependency at two timepoints (e.g., T1 vs T2) to infer the presence of holistic forgetting (when no evidence for a decrease in dependency is seen) or fragmented forgetting (where a significance decrease in dependency is seen between timepoints).

2.1.5.3. Preregistered analyses. We preregistered six statistical contrasts, three relating to retrieval accuracy and three to retrieval dependency. In relation to retrieval accuracy, we first used one-sample *t*-tests (alpha $p < .05$) comparing retrieval accuracy at T1 and T2 separately to 0.25 (i.e., chance level, given 4AFC test). We next directly compared retrieval accuracy in the immediate relative to delayed condition in a paired-sample *t*-test to assess forgetting. Similar one-sample *t*-tests were conducted for retrieval dependency (comparing to 0) to assess dependency at both T1 and T2, and T1 and T2 were directly compared using a paired-sample t-test to assess dependency as a function of delay. We also report Bayes factors for all contrasts using Bayesian t-tests with a prior Cauchy distribution ($r = 0.707$) centred at 0. When evidence favours the alternative hypothesis we present BF10 and when evidence favours the null hypothesis we present BF01. All statistical analyses were conducted in JASP (JASP Team (2022) JASP (Version 0.16.3)). We did not preregister any analyses related to confidence (apart from in relation to correlations with VVIQ, which will be detailed in a separate publication).

2.1.6. Transparency and openness

In all studies we report how we have determined sample size, how we set data collection stopping rules (in relation to Bayes Factors), all data exclusions, all manipulations, and all measures collected. All preregistrations, data, and statistical analyses are available on the Open Science Framework (https://osf.io/zgra9/). Initial data analysis was conducted using MATLAB and R, including the calculation of retrieval dependency in R. All statistical analyses were conducted using JASP and R. All data figures were created in MATLAB.

2.2. Results

2.2.1. Preregistered analyses

Means (and SDs) for retrieval accuracy per condition across all experiments are reported in Table 2. Means (and SDs) for raw proportion of joint retrievals for the observed data and independent model, and retrieval dependency, are reported in Table 3. We also report means (and SDs) for confidence in Table 4, however these are not analysed further in this publication.

Participants remembered the episodic events well at T1 (Mean accuracy = 0.77, SD = 0.17); however, they experienced forgetting over the 3-day delay (T2 Mean accuracy = 0.42, SD = 0.14; see Fig. 2). Retrieval accuracy was significantly above chance (i.e., 0.25) at both T1 ($t_{38} = 19.32$, $p < .001$, $d = 3.09$, $BF_{10} = 1.35 \times 10^{18}$) and T2 ($t_{38} = 7.85$, $p < .001$, $d = 1.26$, $BF_{10} = 6.59 \times 10^{6}$). There was a significant reduction in retrieval accuracy from T1 to T2 ($t_{38} = 15.43$, $p < .001$, $d = 2.47$, $BF_{10} = 8.86 \times 10^{14}$), providing clear evidence of forgetting.

Table 2

Means (and Standard Deviations) for retrieval accuracy across Experiments 1–5 at T1, and T2 (and T3 for Experiment 4).

	T1	T2	T3
Experiment 1 – Events Between-subject	0.77 (0.17)	0.42 (0.14)	
Experiment 2 – Objects Between-subject	0.59 (0.15)	0.37 (0.05)	
Experiment 3 – Events Within-subject	0.73 (0.17)	0.49 (0.13)	
Experiment 3 – Objects Within-subject	0.55 (0.13)	0.47 (0.10)	
Experiment 4 – Objects	0.66 (0.13)	0.55 (0.13)	0.44 (0.10)
Experiment 5 – Objects Short Encoding	0.58 (0.19)	0.45 (0.20)	
Experiment 5 – Objects Long Encoding	0.67 (0.16)	0.58 (0.15)	

We found clear evidence for retrieval dependency at both T1 ($t_{38} = 5.55$, $p < .001$, $d = 0.89$, $BF_{10} = 7749$) and T2 ($t_{38} = 4.67$, $p < .001$, $d = 0.75$, $BF_{10} = 602$). Importantly, there was no evidence for a decrease in retrieval dependency across delay ($t_{38} = 1.65$, $p = .108$, $d = 0.26$, $BF_{01} = 1.68$). Although the evidence for the null was anecdotal, the Bayes factor indicated more support for the null (i.e., no change in dependency) than the alternative hypothesis. Note that several participants (N = 6) reached floor performance at T2 (retrieval accuracy <30 %), which may have contributed to a numerical decrease in average retrieval dependency.

2.2.2. Non-preregistered analyses

Given the relatively high number of participants ($N = 6$) with <0.30 accuracy at T2, we performed the same statistical analyses as above but with the removal of these low performing participants. Retrieval accuracy was significantly above chance (i.e., 0.25) at both T1 ($t_{32} = 21.98$, $p < .001$, $d = 3.83$, $BF_{10} = 5.49 \times 10^{17}$) and T2 ($t_{32} = 9.10$, $p < .001$, $d = 1.58$, $BF_{10} = 4.81 \times 10^{7}$). We observed forgetting of episodic events from T1 (Mean accuracy = 0.80, SD = 0.14) to T2 (Mean accuracy = 0.45, SD = 0.13) ($t_{32} = 15.23$, $p < .001$, $d = 2.65$, $BF_{10} = 1.79 \times 10^{13}$). As previously with the full sample, we observed retrieval dependency at both T1 ($t_{32} = 5.70$, $p < .001$, $d = 0.99$, $BF_{10} = 7240$) and T2 ($t_{32} = 4.65$, $p < .001$, $d = 0.81$, $BF_{10} = 439$). There was no evidence for a decrease in retrieval dependency across delay ($t_{32} = 1.31$, $p = .200$, $d = 0.23$, $BF_{01} = 2.46$), with more evidence in favour of the null (i.e., no evidence of a decrease in dependency) than alternative hypothesis.

2.3. Discussion

The results of Experiment 1 suggested that episodic events learnt by viewing images and imagining associations between separate elements (i.e., an object, a location, and a famous person) are retrieved and forgotten holistically, as previously observed (Horner et al., 2015; Joensen et al., 2020). That is, if one element of an episodic event is retrieved, the other elements tend to be retrieved together but if the event is forgotten, it tends to be forgotten in its entirety. This statistical dependency between the retrieval of event elements did not decrease over time. These findings extend our previous findings by showing that this pattern of results is also seen for visual stimuli (images of locations, people, and objects) rather than verbal stimuli (written words), increasing the generalisability of our previous results. Given this replication and extension our next aim was to determine whether holistic retrieval and forgetting is observed for objects.

3. Experiment 2

In Experiment 2 we asked whether object features (specifically, object state and colour) are retrieved together and if they fragment with forgetting (as in Brady et al., 2013). We used the same episodic memory task where participants encoded event associations, but tested memory

Table 3
Means (and Standard Deviations) for joint (non)retrievals for the data and independent model and retrieval dependency (joint retrievals data – joint retrievals independent model) across Experiments (E) 1–5 at T1, and T2 (and T3 for Experiment 4).

	Data			Independent Model			Retrieval Dependency		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
E1 – Events	0.74 (0.13)	0.57 (0.06)		0.70 (0.14)	0.55 (0.05)		0.04 (0.05)	0.03 (0.04)	
E2 – Objects	0.71 (0.10)	0.55 (0.06)		0.63 (0.10)	0.52 (0.02)		0.08 (0.05)	0.03 (0.06)	
E3 – Events	0.69 (0.11)	0.56 (0.07)		0.65 (0.13)	0.52 (0.03)		0.04 (0.04)	0.04 (0.06)	
E3 – Objects	0.67 (0.10)	0.62 (0.11)		0.61 (0.10)	0.55 (0.04)		0.07 (0.08)	0.07 (0.10)	
E4 – Objects	0.74 (0.11)	0.66 (0.10)	0.61 (0.09)	0.68 (0.09)	0.61 (0.08)	0.54 (0.05)	0.07 (0.07)	0.05 (0.07)	0.07 (0.08)
E5 – Objects Short Encoding	0.73 (0.15)	0.59 (0.18)		0.62 (0.12)	0.57 (0.11)		0.11 (0.13)	0.02 (0.14)	
E5 – Objects Long Encoding	0.76 (0.13)	0.69 (0.13)		0.68 (0.12)	0.61 (0.11)		0.08 (0.10)	0.08 (0.11)	

Table 4
Means (and Standard Deviations) for retrieval confidence across Experiments 1–5 at T1, and T2 (and T3 for Experiment 4).

	T1	T2	T3
Experiment 1 – Events Between-subject	2.85 (0.67)	1.60 (0.55)	
Experiment 2 – Objects Between-subject	2.64 (0.45)	1.80 (0.46)	
Experiment 3 – Events Within-subject	2.76 (0.52)	1.79 (0.47)	
Experiment 3 – Objects Within-subject	2.61 (0.48)	2.02 (0.45)	
Experiment 4 – Objects	2.94 (0.41)	2.46 (0.48)	1.97 (0.54)
Experiment 5 – Objects Short Encoding	2.77 (0.52)	2.09 (0.61)	
Experiment 5 – Objects Long Encoding	3.01 (0.48)	2.56 (0.55)	

for the object features. We hypothesised that objects should follow an independent forgetting trajectory. Although retrieval of object features may show retrieval dependency immediately after encoding, dependency should decrease over time.

Experiment 2 used the same experimental structure as Experiment 1, with participants learning three-element events. However, we tested their memory for the features (state and colour) of the objects at T1 and T2. This is a conceptual replication of Brady et al. (2013) with the difference between the studies being that in Experiment 2 objects were encoded as a part of three-element episodic events rather than as single objects. We used a retrieval task that was as close as possible to the original Brady et al. (2013) experiments. Participants were asked to select the object with the correct colour and state among foils of the same object that had either only the correct colour or state, or neither of the correct features.

3.1. Methods

Experiment 2 was identical to Experiment 1 with the following exceptions.

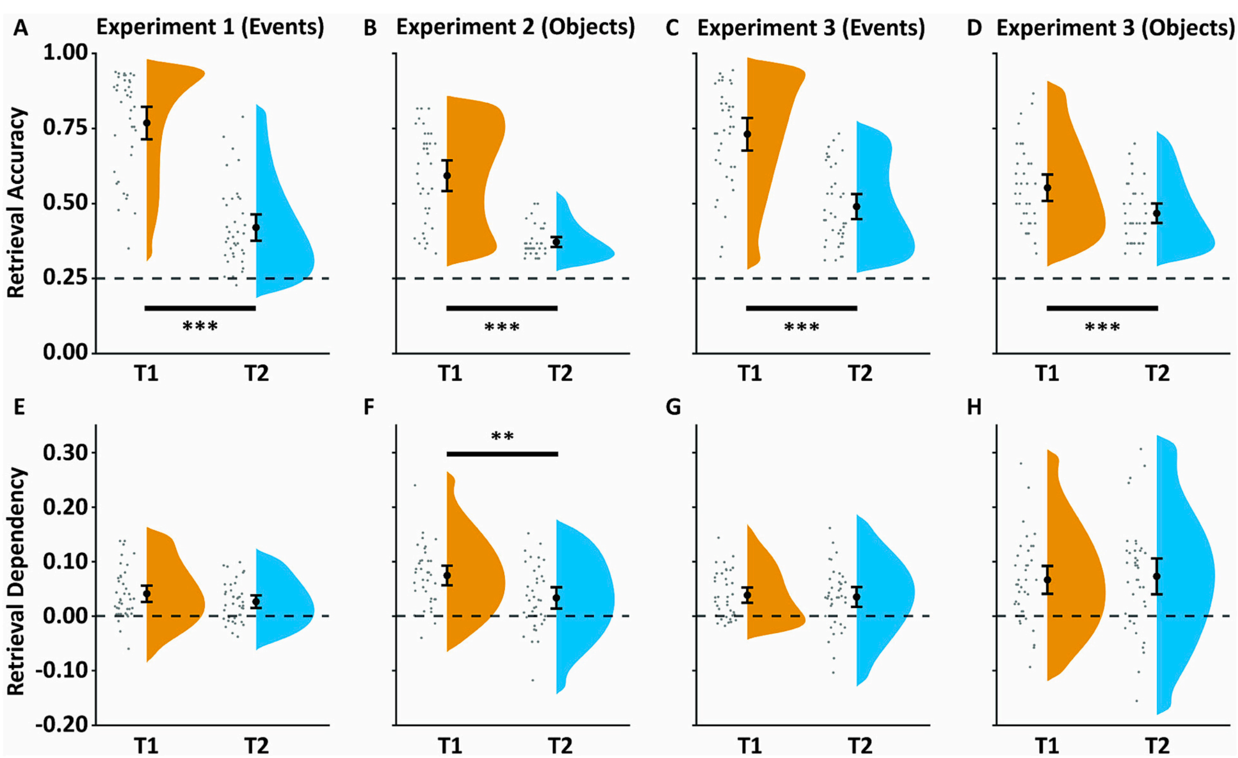

Fig. 2. Retrieval accuracy and dependency across experiments 1–3.
Retrieval accuracy (top row) and retrieval dependency (bottom row) at T1 (immediate test) and T2 (delayed test) across Experiments 1–2, with each plot showing individual participant means (grey dots), mean and 95 % confidence intervals (black circle and error bars), and distribution of sample (orange and blue 'rainclouds' showing kernel density estimates). Retrieval accuracy for (A) Experiment 1 (Events), (B) Experiment 2 (Objects), (C) Experiment 3 (Events), and (D) Experiment 3 (Objects). Retrieval dependency for (E) Experiment 1 (Events), (F) Experiment 2 (Objects), (G) Experiment 3 (Events), and (H) Experiment 3 (Objects). Bars with asterisks show comparisons between T1 and T2 that are significant; ** $p < .01$, *** $p < .001$. Raincloud plots created using customised MATLAB scripts using core code elements from Allen et al. (2021).

3.1.1. Participants

Seventy-nine participants were recruited for this two-part online experiment via Prolific (www.prolific.co). To be eligible for recruitment, participants had to be native English speakers (aged 18 to 40) who have lived in the UK for 5 or more years, with normal or corrected-to-normal vision and no colourblindness. Of the eligible participants who started the experiment, 37 participants (Mean age = 29.73, SD = 7.17; 10 male, 27 female) completed both sessions, met the pre-registered inclusion criteria and were retained in the analyses. The other 42 participants (53 %) were excluded due to one of the following reasons: not returning for the final session (13 participants), having an immediate retrieval accuracy below 30 % or with more than 9 successive timeouts (6), or having a delay retrieval accuracy at or below 30 % (23). As in Experiment 1, we excluded based on retrieval accuracy however our theoretical conclusions are primarily drawn from the independent measure of retrieval dependency.

The sample size calculation was the same as reported for Experiment 1. The pre-registered inclusion criteria were that retrieval accuracy should be above 30 % and below 95 % at T1 and T2. Additionally, participants needed to have completed at least one of four attention checks during the study phase at T1, and not have more than 9 successive timeout responses at retrieval. The attention checks involved randomly intermixed trials requesting participants to press the spacebar within 3 s during the presentation of the to-be-learned events. The exclusion criteria were different from Experiment 1; in this experiment we excluded participants above ceiling and below chance performance at both T1 and T2 (rather than just T1 as in Experiment 1) to correctly estimate retrieval dependency at both timepoints given the increased difficulty of retrieving object features relative to events. For the full preregistration of this experiment, see: https://osf.io/gxdtr.

3.1.2. Procedure

On each encoding trial, participants saw a fixation point on the screen for 500 ms, followed by three images for 5 s, followed by a blank screen with the instruction "Imagine…" for 4 s. The encoding trials in the present study were longer (5 s + 4 s = 9 s) than in Experiment 1 (2.5 s + 4 s = 6.5 s) because after piloting the experiment we found that many participants were at chance performance (25 %) with the original timings, owing to the fact that object features are harder to recall than event elements. We increased the study timings accordingly in the preregistration of Experiment 2, to avoid floor performance and to be able to estimate retrieval dependency at both time points. There was a blank inter-trial interval of 500 ms before the onset of the next trial.

In the test phase, participants' memory for the object features was tested by presenting either the location or person from an event as a cue (e.g., hair salon) and asking them to select the associated target object in the correct state and colour (e.g., a blue cup with tea). The target was presented among three foils from the same category with combinations of correct and incorrect features (i.e., a blue empty cup, a yellow cup with tea, a yellow empty cup). Participants were tested with each of the two possible cues from the same event (i.e., cue location and retrieve object, cue person and retrieve object) in two separate blocks and their performance was averaged across the blocks. The test phase started with on-screen instructions and one practice trial. Half of the objects were tested at T1 (60 trials) and half were tested at T2 (60 trials). Within each test session, trials were completed in two blocks of 30 trials, with an opportunity to take a break in between the blocks. On each trial, participants saw a fixation point on the screen for 500 ms, then the cue element image appeared alone for 2 s, followed by the four response option images appearing below the cue. The order of the response options on the screen from left to right was random. The options remained on the screen for 5 s or until participants made a response by pressing one of the four keys (1–4). The response time was increased from 4 s to 5 s in the present experiment, relative to Experiment 1, as piloting indicated people took longer to recall object features. After a 500 ms blank interval participants saw a prompt "How confident are you in your choice?" on the screen and were asked to press the keys 1–4 to rate their confidence. The confidence rating prompt remained on the screen for 4 s or until response. There was a blank inter-trial interval of 500 ms before the onset of the next trial.

After completing the test phase of the delay session, participants completed the VVIQ and were asked to report whether they had prior knowledge of the famous people in the same celebrity check task used in Experiment 1. Participants reported knowing 89 % (mean across participants; SD = 11 %) of the famous people that were presented.

3.1.3. Analyses

3.1.3.1. Retrieval accuracy and confidence.
Mean retrieval accuracy and mean retrieval confidence were calculated for T1 and T2. As detailed in the preregistration, timeouts on memory trials were converted to incorrect responses for the purposes of calculating retrieval accuracy and dependency. Overall, 2.57 % (SD = 2.24 %) of the responses were timeouts at T1 and 1.67 % (SD = 1.80) were timeouts at T2.

3.1.3.2. Retrieval dependency.
We created two contingency tables per participant for T1 and T2 separately. One table was the count of (un) successful retrievals of the object state and colour when cued by the location, and the other was when cued by the person. We created equivalent contingency tables for the independent model as in Experiment 1. We then calculated the proportion of joint retrievals for each table and took the mean of this measure across the two separate cues (location and person). We then calculated retrieval dependency as the difference between joint retrievals in the observed data and independent model, resulting in a single measure of retrieval dependency at both T1 and T2.

3.1.3.3. Preregistered analyses.
The preregistered statistical contrasts were identical to those of Experiment 1.

3.2. Results

3.2.1. Preregistered analyses

Performance was relatively high for remembering both the state and colour (combined) at T1 (Mean = 0.59, SD = 0.15) and was lower at T2 (Mean = 0.37, SD = 0.05; see Fig. 2). Retrieval accuracy was significantly above chance level (i.e., 25 %) at T1 ($t_{36} = 13.62$, $p < .001$, $d = 2.24$, $BF_{10} = 6.85 \times 10^{12}$) and T2 ($t_{36} = 14.89$, $p < .001$, $d = 2.45$, $BF_{10} = 9.72 \times 10^{13}$). There was a significant reduction in retrieval accuracy from T1 to T2 ($t_{36} = 9.16$, $p < .001$, $d = 1.51$, $BF_{10} = 1.60 \times 10^{8}$), providing clear evidence of forgetting.

As in Experiment 1, there was significant retrieval dependency at T1 ($t_{36} = 8.36$, $p < .001$, $d = 1.38$, $BF_{10} = 1.88 \times 10^{7}$) and T2 ($t_{36} = 3.48$, $p = .001$, $d = 0.57$, $BF_{10} = 24$). However, in contrast to Experiment 1, there was also a significant decrease in retrieval dependency with delay ($t_{36} = 3.05$, $p = .004$, $d = 0.50$, $BF_{10} = 8.64$), with strong evidence in favour of the alternative hypothesis.

3.2.2. Non-preregistered analyses

Given evidence for fragmentation in Experiment 2, we asked whether this was associated with initial performance across participants. It could be that fragmentation for object features is simply driven by poor encoding. If this was the case, we would expect to see a correlation between initial performance at T1 and the difference in dependency between T1 and T2, presuming initial T1 performance is related to the strength of initial encoding. We saw no evidence for a significant correlation between these measures ($r_{35} = 0.03$, $p = .839$, 95 % CI [−0.29, 0.35]), suggesting fragmentation was not associated with initial retrieval accuracy across participants.

We next asked whether participants were better at remembering either the state or colour of studied objects. For this, we assessed accuracy with a 2 (Object Feature: State/Colour) x 2 (Delay: T1/T2) Repeated Measures ANOVA. Results showed a significant main effect of Delay only, $F(1,36) = 73.27$, $p < .001$, $\eta_p^2 = 0.671$, whereas the main effect of Object Feature and the interaction were not significant, $F(1,36) = 2.98$, $p = .093$, $\eta_p^2 = 0.076$, $F(1,36) = 0.90$, $p = .349$, $\eta_p^2 = 0.024$, respectively, showing similar forgetting of both state and colour features.

Finally, we directly compared the difference in retrieval dependency between the testing sessions T1 and T2 in Experiment 1 (event memory) with Experiment 2 (object memory). An independent samples *t*-test revealed that the difference in retrieval dependency over the two testing sessions (T1-T2) was not statistically different between Experiment 1 (M = 0.015) and Experiment 2 (M = 0.041) ($t_{74} = 1.67$, $p = .100$, $d = 0.38$, $BF_{01} = 1.28$), showing weak evidence in favour of the null hypothesis.

3.3. Discussion

Experiment 2 provided evidence for a decrease in retrieval dependency over delay, consistent with the proposal that object features are forgotten relatively independently. This contrasts with Experiment 1 where retrieval dependency was consistent over delay. Across the two experiments, we therefore provide evidence for more holistic forgetting of event elements and more fragmented forgetting of object features. However, in a non-preregistered across-experiment comparison there was not a significant difference between experiments in relation to the decrease in retrieval dependency as a function of delay. We next sought to replicate this possible dissociation in a within-subjects manner, testing memory for both event elements and object features in the same participants.

4. Experiment 3

Experiment 3 tested both event and object memory. To ensure retrieval accuracy for both events and objects at T1 and T2 was above floor and not at ceiling, we adjusted both the encoding time (12 s in total) and the time of T1 (1 h after encoding) and T2 (2 days after encoding). We also changed our sampling and analysis approach from a Null Hypothesis Significance Testing (NHST) to a Bayesian approach, where we set a Bayesian stopping rule of BF > 6 (either for the alternative or null hypothesis).

4.1. Method

4.1.1. Participants

One hundred and forty participants were recruited for this two-part online experiment via Prolific (www.prolific.co). To be eligible for recruitment, participants had to be native English speakers (aged 18 to 40) who have lived in the UK for 5 or more years, with normal or corrected-to-normal vision and no colour-blindness. Of the eligible participants who started the experiment, only 38 participants (Mean age = 28.39, SD = 5.84; 10 male, 28 female) completed T1 and T2, met the pre-registered inclusion criteria and were retained in the analyses. The other 102 participants (73 %) were excluded due to one of the following reasons: not returning for the testing sessions (T1 or T2) (17 participants), having retrieval accuracy at or below 30 % for either objects or events at any of the testing sessions (T1 or T2) (47, of those mainly objects at T2), having retrieval accuracy at or above 95 % for either objects or events at any of the testing sessions (31, of those mainly events at T1), or having more timeouts than allowed or missing attention checks (7). The 140 upper limit of the sample size was determined by resource constraints. We return to the high exclusion rate in the discussion of Experiment 3. We return to the high exclusion rate in the discussion, however, again note we are excluding based on retrieval accuracy but our primary independent variable is retrieval dependency.

The data collection plan was to recruit participants for this study until we reached Bayes factors in favour of the alternative hypothesis ($BF_{10} > 6$) or in favour of the null hypothesis ($BF_{01} > 6$) for the preregistered comparisons, or until we reach a maximum sample of 128 participants (useable data-sets, set by budget constraints). The preregistered inclusion criteria for performance were that the retrieval accuracy in the episodic memory task should be above 30 % and below 95 % at T1 and T2. Additionally, participants needed to have passed at least one of four attention checks during the study phase at T1, and not to have 6 or more successive timeout responses at retrieval. These exclusion criteria were largely the same as in Experiment 2, except fewer timeouts (i.e., maximum of 5 rather than 8) were permissible, given fewer trials. For the full preregistration of this experiment, see: https://osf.io/kjq4u.

4.1.2. Design

Experiment 3 had a within-subjects design with a factor of Delay (two levels: Test phase T1 and T2) and Test-type (two levels: event elements and object features). T1 was conducted 1 h after encoding and T2 was conducted 2 days after encoding. The dependent measures of interest for the preregistered analyses were retrieval accuracy (proportion) and retrieval dependency of event elements.

4.1.3. Procedure

This three-part experiment consisted of a single study session, a delayed test session after 1 h (T1), and second delayed test session after 2 days (T2). Each test phase consisted of two blocks testing memory for the event associations and for object features respectively. Participants were recruited online via Prolific and signed informed consent. A demographic questionnaire, detailed on-screen instructions, and one study phase practice trial were presented before participants could proceed to the study phase.

In the study phase, participants learnt 60 novel events. On each trial, participants first saw a fixation point on the screen for 500 ms, followed by three images for 6 s, followed by a blank screen with the instruction “Imagine…” for 6 s. The study phase trials in the present study were longer (6 s + 6 s = 12 s) than those of Experiment 1 (2.5 s + 4 s = 6.5 s) and Experiment 2 (5 s + 4 s = 9 s) to ensure that there would not be too many exclusions from the last testing session due to low performance on the object memory trials. There was a blank inter-trial interval of 500 ms before the onset of the next trial. The study phase consisted of 60 trials, separated into four blocks. Participants were given an opportunity to take a break in between the blocks.

In each test phase, participants’ memory for the object features and event associations was tested in separate blocks. The structure of the event and object memory trials was similar to Experiments 1 and 2 respectively. The 60 encoded events were randomly assigned to the four conditions: events T1, objects T1, events T2, and objects T2, resulting in 15 events per condition. In the object blocks at T1 and T2, the 15 objects were presented twice, once with a location cue and once with a person cue. On each object and event test trial, participants saw a fixation point on the screen for 500 ms, then the cue element image appeared alone for 2 s, followed by the four response option images appearing below the cue. The options remained on the screen for 6 s or until the participant made a response by pressing one of the four keys (1–4). For object trials the response options were the correct object with lures in different combinations of incorrect state and colour. For event trials the response options were the correct event element with lures of the same element category (i.e., object, location, or person) from different events. The order of the event and object trial blocks was counterbalanced between participants. The response time was increased from 5 s to 6 s in the present experiment, relative to Experiment 2, to minimise floor performance and timeouts. After a 500 ms blank interval participants saw a prompt “How confident are you in your choice?” on the screen, and were asked to press the keys 1–4 to rate their confidence. The confidence rating prompt remained on the screen for 6 s or until response. There was a blank inter-trial interval of 500 ms before the onset of the next trial.

After completing T2, participants were asked to complete the VVIQ questionnaire (Marks, 1973) to leave any feedback about the experiment in a comments box after completing each of the sessions and were reimbursed for their time at a rate of £8–9 per hour via Prolific. If participants did not meet the inclusion criteria at T1, they were still reimbursed for their time but were not invited back for T2.

4.1.4. Analyses

4.1.4.1. Retrieval accuracy and confidence. Mean retrieval accuracy and mean retrieval confidence were calculated for events and objects at T1 and T2. As detailed in our preregistration, timeouts on memory trials were converted to incorrect responses for the purposes of calculating retrieval accuracy and dependency. For events, 2.11 % (SD = 3.16 %) of responses were timeouts at T1 and 2.40 % (SD = 5.56 %) were timeouts at T2. For objects, 2.28 % (SD = 4.66 %) of responses were timeouts at T1 and 1.75 % (SD = 3.07 %) were timeouts at T2.

4.1.4.2. Retrieval dependency. Retrieval dependency was calculated in the same manner as Experiment 1 for events and Experiment 2 for objects, but with fewer trials per contingency table (i.e., 15 in Experiment 3, 30 in Experiments 1–2).

4.1.4.3. Preregistered analyses. We preregistered 13 statistical contrasts, six relating to retrieval accuracy and seven to retrieval dependency. In relation to retrieval accuracy, we first used four Bayesian one-sample *t*-tests comparing retrieval accuracy at T1 and T2 for objects and events separately to 0.25 (i.e., chance level). We next directly compared retrieval accuracy between T1 and T2 in a Bayesian paired-sample t-test to assess for forgetting of objects and events separately.

Similar Bayesian one-sample t-tests were conducted for retrieval dependency (compared to 0) to assess dependency at T1 and T2 for objects and events. We next compared retrieval dependency between T1 and T2, to assess for potential decreases in dependency over delay, for objects and events separately in Bayesian paired-sample t-tests. Finally, to assess for differences in dependency over time for objects and events, we directly compared the difference in T1 vs T2 dependency for objects and events in a Bayesian paired-sample t-test. All Bayesian t-tests used a prior Cauchy distribution ($r = 0.707$) centred at 0. When evidence favours the alternative hypothesis we present BF10 and when evidence favours the null hypothesis we present BF01. We also report more standard NHST statistics for completeness.

4.2. Results

4.2.1. Preregistered analyses

Participants' memory for the event associations (e.g., person-location) was high at T1 (Mean accuracy = 0.73, SD = 0.16); however, they experienced forgetting over the 2-day delay (Mean accuracy = 0.49, SD = 0.13; see Fig. 2). Retrieval accuracy was above chance (i.e., 0.25) at both T1 ($BF_{10} = 6.66 \times 10^{16}$, $t_{37} = 18.01$, $p < .001$, $d = 2.92$) and T2 ($BF_{10} = 1.01 \times 10^{11}$, $t_{37} = 11.58$, $p < .001$, $d = 1.88$). There was a reduction in retrieval accuracy from T1 to T2 ($BF_{10} = 7.16 \times 10^{8}$, $t_{37} = 9.63$, $p < .001$, $d = 1.56$), providing clear evidence of forgetting.

Participants remembered the object features (i.e., state and colour combined) at T1 (Mean accuracy for both features combined = 0.55, SD = 0.13) and T2 (Mean accuracy = 0.47, SD = 0.10). Retrieval accuracy was above chance (i.e., 0.25) at both T1 ($BF_{10} = 1.99 \times 10^{13}$, $t_{37} = 13.88$, $p < .001$, $d = 2.25$) and T2 ($BF_{10} = 1.01 \times 10^{13}$, $t_{37} = 13.58$, $p < .001$, $d = 2.20$). There was a reduction in retrieval accuracy from T1 to T2 ($BF_{10} = 22$, $t_{37} = 3.44$, $p = .001$, $d = 0.56$), providing clear evidence of forgetting.

Evidence for retrieval dependency was seen for events at both T1 ($BF_{10} = 7586$, $t_{37} = 5.57$, $p < .001$, $d = 0.90$) and T2 ($BF_{10} = 81$, $t_{37} = 3.95$, $p < .001$, $d = 0.64$). As in Experiment 1, no evidence for a decrease in dependency between T1 and T2 was seen ($BF_{01} = 5.52$, $t_{37} = 0.28$, $p = .779$, $d = 0.05$), with more evidence in favour of the null hypothesis (no change in dependency across delay). A similar pattern was seen for object features. We saw evidence for retrieval dependency at T1 ($BF_{10} = 3329$, $t_{37} = 5.28$, p < .001, $d = 0.86$) and T2 ($BF_{10} = 372$, $t_{37} = 4.51$, $p < .001$, $d = 0.73$). Contrary to our hypothesis and the results of Experiment 2, we saw no evidence for a decrease in dependency for objects features over delay ($BF_{01} = 5.52$, $t_{37} = -0.28$, $p = .781$, $d = 0.05$), with more evidence in favour of the null hypothesis. Finally, a direct comparison between the difference in retrieval dependency between T1 and T2 for objects and events provided support for the null hypothesis (i.e., no difference in the change in dependency over delay between objects and events) ($BF_{01} = 5.22$, $t_{37} = 0.44$, $p = .660$, $d = 0.07$).

4.2.2. Non-preregistered analyses

Although unlikely to bias our results in relation to changes in dependency across delay, we were concerned with the high exclusion rate in Experiment 3. This is a result of having identical encoding conditions while ensuring event performance was off ceiling at T1 and object performance was off floor at T2. We therefore reanalysed the event and object data separately, excluding participants solely on the basis of event and object performance respectively. This decreased the exclusion rate to 59 % for events and 51 % for objects. Despite the increased number of participants in both tests, our results were consistent with the preregistered analyses. There was not a significance decrease in dependency between T1 and T2 for either events ($t_{57} = 0.58$, $p = .563$, $d = 0.08$, $BF_{01} = 5.93$) or objects ($t_{68} = -0.28$, $p = .782$, $d = 0.03$, $BF_{01} = 7.29$).

As in Experiment 2, in Experiment 3 we checked that participants' object memory did not differ according to the feature dimension that participants were set to remember, i.e., the objects' state or colour. The 2 × 2 Repeated Measures ANOVA with factors of Object Feature (State/Colour) and Delay (T1/T2) on participants' accuracy showed a significant main effect of Delay only, $F(1,37) = 11.33$, $p = .002$, $\eta_p^2 = 0.234$, and no significant main effect of Object Feature, $F(1,37) = 0.85$, $p = .363$, $\eta_p^2 = 0.022$, nor an interaction, $F(1,37) = 1.84$, $p = .183$, $\eta_p^2 = 0.047$. Once again, significant forgetting was found irrespective of feature dimension.

For completeness, we also report a set of analyses that assessed participants' accuracy and retrieval dependency for event and object information in the two test delays with omnibus Repeated Measures ANOVAs. These had factors of Test Type (Event/Object) and Delay (T1/T2). The ANOVA on accuracy showed significant main effects of Test Type, $F(1, 37) = 22.14$, $p < .001$, $\eta_p^2 = 0.374$, and Delay, $F(1, 37) = 63.11$, p < .001, $\eta_p^2 = 0.630$, qualified by a significant interaction, $F(1, 37) = 30.63$, $p < .001$, $\eta_p^2 = 0.453$. Follow-up pairwise *t*-tests showed that while accuracy for both events and object features decreased between T1 and T2, accuracy for event information decreased faster, such that participants remembered more event information than object information at T1 ($t_{37} = 6.13$, $p > .001$, $d = 0.99$, $BF_{10} = 38{,}275$) but not in the delayed test, where accuracy did not significantly differ ($t_{37} = 1.04$, $p = .306$, $d = 0.17$, $BF_{01} = 3.48$). The ANOVA run on retrieval dependency instead only revealed a significant main effect of Test Type, $F(1, 37) = 8.86$, $p = .005$, $\eta_p^2 = 0.193$, and a non-significant main effect of Delay, $F(1, 37) = 0.01$, $p = .919$, $\eta_p^2 < 0.001$, and interaction, $F(1, 37) = 0.21$, $p = .653$, $\eta_p^2 = 0.006$. This shows that participants exhibited greater retrieval dependency for object than event information, overall.

Finally, we assessed the degree of fragmentation for object features across participants as a function of retrieval accuracy at T1. As in Experiment 2, this showed no evidence for an association ($r_{36} = -0.14$, $p = .418$, 95 % CI [−0.44, 0.19]). Nor was any evidence for an association seen when collapsing participants across Experiments 2 and 3 for object features ($r_{73} = -0.04$, $p = .760$, 95 % CI [−0.26, 0.19]). Thus, initial encoding strength is unlikely to be a driver of fragmentation for object features.

4.3. Discussion

Experiment 3 assessed memory for both objects and events after a 1-h and 2-day delay. We predicted retrieval dependency would remain consistent across delay for events (as in Experiment 1) but decrease for object features (as in Experiment 2). We replicated the lack of decrease in dependency for events seen in Experiment 1, providing evidence for the null that no change in dependency occurs. However, we failed to replicate the decrease in dependency seen in Experiment 2 for object features. Indeed, we provide support for the null hypothesis that no change in dependency is seen across delay.

The exclusion rate in Experiment 3 (and to a lesser extent Experiments 1–2) was very high. This was a function of needing to ensure we could accurately assess retrieval dependency for event elements and object features at both T1 and T2 in each participant. Most participants were excluded due to either ceiling effects for events at T1 or floor effects for objects at T2. Although our measure of retrieval dependency controls for overall retrieval accuracy, it is still possible (though highly unlikely) that our exclusions have somehow biased our ability to reveal potential decreases in retrieval dependency across delay. As such, we analysed events and objects independently, including as many participants as possible within each separate analysis. These analyses decreased the exclusion rate from 73 % to 59 % for Events and 51 % for Objects. These exploratory analyses revealed the same pattern of results, with no evidence for a decrease in dependency between T1 and T2 for both event elements and object features. Thus, our high exclusion rate based on retrieval accuracy is unlikely to have affected our primary results in relation to retrieval dependency.

Before discussing differences between Experiment 2 and 3, one consistent finding was that forgetting for object state and colour did not differ. Thus, the presence or absence of object fragmentation does not appear to be dependent on differential forgetting rates for object state and colour. There were several differences between Experiment 2 (where we saw a decrease in dependency for object features across delay) and Experiment 3 (where we saw no evidence for a decrease): (1) encoding time per trial was shorter in Experiment 2 (5 s + 4 s) than Experiment 3 (6 s + 6 s), (2) the delay between encoding and T1 was shorter in Experiment 2 (immediate) than Experiment 3 (1 h), and (3) the delay between encoding and T2 was longer in Experiment 2 (3 days) relative to Experiment 3 (2 days). Interestingly, although forgetting was seen between T1 and T2 in Experiment 3 (decreasing from 0.55 to 0.47), this is numerically less forgetting than seen in Experiment 2 (decreasing from 0.59 to 0.37). It is possible that the short 1-h delay between encoding and T1 allowed for both forgetting and fragmentation during this period, which we may have missed. The shorter delay between encoding and T2 may have further decreased the likelihood of seeing fragmentation of object features. Given this possibility, in Experiment 4 we tested memory for object features at three time points: immediately after encoding, 5 h after encoding, and 2 days after encoding. We predicted that, if fragmentation for object features occurs, it should be seen particularly when comparing the immediate and 2-day delay conditions.

A further possibility is that the testing of both objects and events (albeit across separate blocks) resulted in participants using differing retrieval strategies and/or retrieval modes, and this could have led to the lack of a decrease in dependency for objects in Experiment 3 relative to Experiment 2. In Experiment 4 we only tested memory for object features, making any 'retrieval mode' more similar to Experiment 2 than 3. We also did not test event memory further due to the consistent evidence for a lack of decrease in dependency for events. Finally, by only testing object features in Experiment 4 we were also able to reduce the exclusion rate relative to Experiment 3.

5. Experiment 4

Experiment 4 tested memory for object features at three time points: immediately after encoding (T1), 5 h after encoding (T2) and 2 days after encoding (T3). We used the same Bayesian analysis approach as in Experiment 3.

5.1. Method

5.1.1. Participants

One hundred and fourteen participants were recruited for this three-part online experiment via Prolific (www.prolific.co). Of the eligible participants who started the experiment, 52 participants (Mean age = 28.15, SD = 6.83; 41 female, 9 male, 2 non-binary) completed all three sessions, met the pre-registered inclusion criteria and were retained in the analyses. Participants were excluded based on the eligibility criteria outlined in our preregistration (see https://osf.io/32zwm). If participants had a retrieval accuracy at or below 30 %, or at or above 95 %, in any of the three testing sessions (T1, T2, T3), they were excluded from further analyses. Participants were also excluded if there were 6 or more successive timeouts in their memory responses. Of the 62 participants (54 %) who were excluded from analyses, 24 did not return for either of the follow-up testing sessions, 20 had retrieval accuracy <30 % at T3, eight had retrieval accuracy <30 % at T2, seven had retrieval accuracy <30 % at T1, and 3 had too many timeout responses. Of the 52 remaining participants, there was a balanced number of participants (11, 13 or 14 in version 1, version 2 and 4, and version 3, respectively) in each of the counterbalanced study lists.

5.1.2. Design

Experiment 3 had a within-subjects design with a factor of Delay (three levels: Test phase T1, T2 and T3). The dependent measures of interest for the preregistered analyses were retrieval accuracy and retrieval dependency of object features.

5.1.3. Procedure

This three-part experiment consisted of an encoding session and immediate test session (T1), a test session after 5 h (T2) and a final test session after 2 days (T3). The study phase in this experiment was the same as in Experiment 3, with the same trial timings. The study phase consisted of 60 trials, separated into four blocks. Participants were given an opportunity to take a break in between the blocks.

In the test phases, the target was presented among three foils from the same category with combinations of correct and incorrect features. Participants were tested with each of the two possible cues from the same event (i.e., cue location and retrieve object, cue person and retrieve object) in two separate blocks and their performance was averaged across the blocks. The test phase started with on-screen instructions and one practice trial. One third of the objects were at T1 (20 events, 40 trials) another third at T2, and another third at T3. Within each session, the test trials were completed in two blocks of 20 trials (i.e., the same objects tested once with location and once with person cues), with an opportunity to take a break in between the blocks. On each trial, participants saw a fixation point on the screen for 500 ms, then the cue element image appeared alone for 2 s, followed by the four response option images appearing below the cue. The options remained on the screen for 6 s or until participants made a response by pressing one of the four keys (1–4). There was a blank inter-trial interval of 500 ms, a confidence rating for 4 s, and a further blank screen for 500 ms before the onset of the next trial. Following the final Test phase participants filled out the VVIQ and the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova, Kozhevnikov, & Motes, 2006). If participants did not meet the inclusion criteria after the immediate session, they were still reimbursed for their time but were not invited back for the delay sessions.

5.1.4. Analyses

5.1.4.1. Retrieval accuracy and confidence. Mean retrieval accuracy and mean retrieval confidence were calculated for test phases T1-T3. As detailed in the preregistration, timeouts on memory trials were converted to incorrect responses for the purposes of calculating retrieval accuracy and dependency. Overall, 2.60 % (SD = 1.98 %) of the responses were timeouts at T1, 1.78 % (SD = 2.54 %) were timeouts at T2, and 2.36 % (SD = 2.59 %) were timeouts at T3.

5.1.4.2. Retrieval dependency. Retrieval dependency was calculated in the same manner as Experiment 2 for objects for T1-T3 separately.

5.1.4.3. Preregistered analyses. We preregistered 11 statistical contrasts, five relating to retrieval accuracy and six to retrieval dependency. In relation to retrieval accuracy, we first used three Bayesian one-sample *t*-tests comparing retrieval accuracy at T1, T2, and T3 to 0.25 (i.e., chance level). We next directly compared retrieval accuracy between T1 and T2 and T1 and T3 in Bayesian paired-sample t-tests to assess for forgetting for object features. Note, our preregistration stated we would "compare accuracy at T1 and T3" however we listed T1 and T2 as the independent variables of interest in the preregistration. We therefore report both comparisons (T1 vs T2 and T1 vs T3) for completeness.

Similar Bayesian one-sample t-tests were conducted for retrieval dependency (comparing to 0) to assess for evidence of dependency at T1, T2, and T3. We next compared retrieval dependency between T1 and T3, T1 and T2, and T2 and T3, to assess for potential decreases in dependency over delay, using Bayesian paired-sample t-tests. Note, we only applied the Bayesian stopping rule to the T1 vs T3 contrast (and listed the other two contrasts as "exploratory" in our preregistration).

5.2. Results

5.2.1. Preregistered analyses

Participants remembered the object features (i.e., state and colour combined) at T1 (Mean accuracy = 0.66, SD = 0.13), T2, (Mean accuracy = 0.55, SD = 0.13), and T3 (Mean accuracy = 0.44, SD = 0.10; see Fig. 3). Retrieval accuracy was above chance at T1 ($BF_{10} = 8.53 \times 10^{24}$, $t_{51} = 22.54$, $p < .001$, $d = 3.13$),

T2 ($BF_{10} = 5.40 \times 10^{18}$, $t_{51} = 16.26$, $p < .001$, $d = 2.26$), and T3 ($BF_{10} = 3.55 \times 10^{15}$, $t_{51} = 13.57$, $p < .001$, $d = 1.88$). There was a significant reduction in retrieval accuracy from T1 to T2 ($BF_{10} = 434{,}932$, $t_{51} = 6.53$, $p < .001$, $d = 0.91$), and from T1 to T3 ($BF_{10} = 1.02 \times 10^{13}$, $t_{51} = 11.61$, $p < .001$, $d = 1.61$), providing clear evidence of forgetting. A further exploratory comparison between retrieval accuracy at T2 and T3 also revealed evidence of forgetting ($BF_{10} = 39{,}010$, $t_{51} = 5.82$, $p < .001$, $d = 0.81$).

Evidence for retrieval dependency was seen at T1 ($BF_{10} = 3.34 \times 10^{6}$, $t_{51} = 7.12$, $p < .001$, $d = 0.99$), T2 ($BF_{10} = 4230$, $t_{51} = 5.15$, $p < .001$, $d = 0.71$), and T3 ($BF_{10} = 200{,}855$, $t_{51} = 6.30$, $p < .001$, $d = 0.87$). Comparing dependency between the three test sessions, we saw evidence in favour of no decrease between T1 and T3 ($BF_{01} = 6.58$, $t_{51} = 0.11$, $p = .916$, $d = 0.02$), T1 and T2 ($BF_{01} = 2.47$, $t_{51} = 1.45$, $p = .153$, $d = 0.20$), and T2 and T3 ($BF_{01} = 3.38$, $t_{51} = -1.20$, $p = .237$, $d = -0.17$).

5.2.2. Non-preregistered analyses

As in Experiment 3, the exclusion rate was high at 54 %. We re-ran the T1 vs T2 contrast excluding only participants who were at floor or

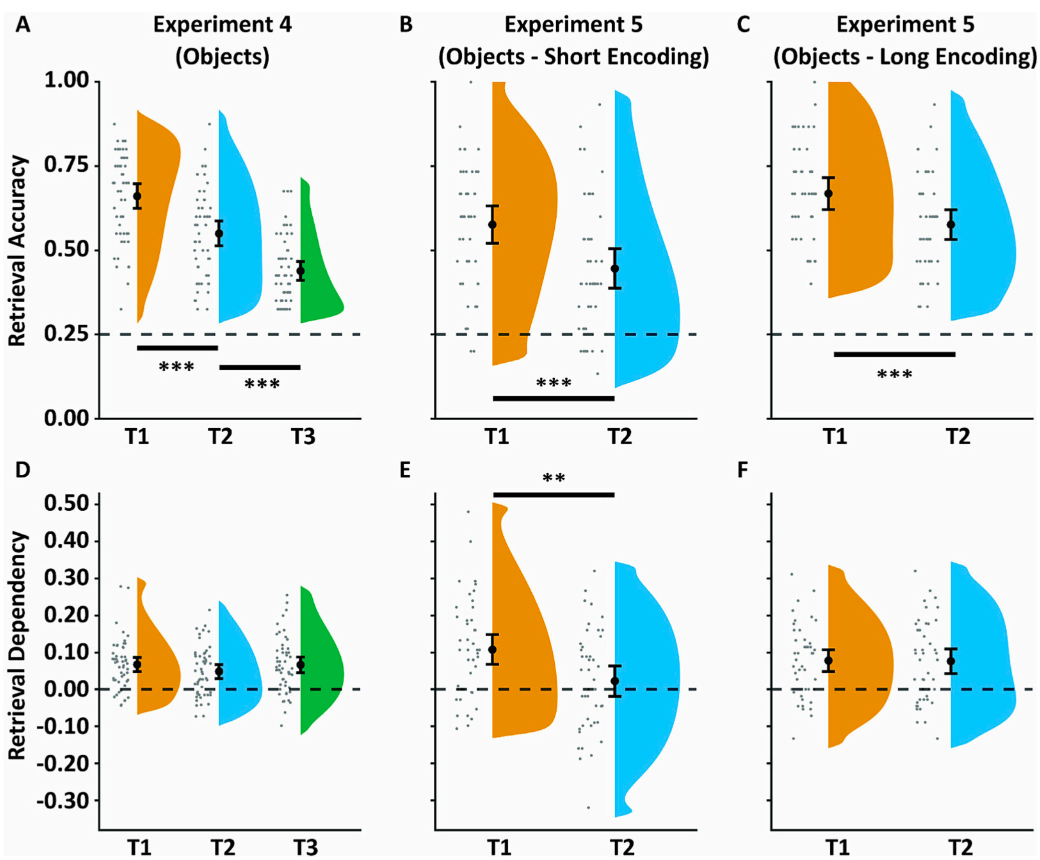

Fig. 3. Retrieval accuracy and dependency across experiments 4–5.
Retrieval accuracy (top row) and retrieval dependency (bottom row) at T1 (immediate test), T2 (delayed test), and T3 (delayed test in Experiment 4 only) across Experiments 4–5, with each plot showing individual participant means (grey dots), mean and 95 % confidence intervals (block circle and error bars), and distribution of sample (orange, blue, and green 'rainclouds' showing kernel density estimates). Retrieval accuracy for (A) Experiment 4 (Objects), (B) Experiment 5 (Objects – Short encoding condition), and (C) Experiment 5 (Objects – Long encoding condition). Retrieval dependency for (D) Experiment 4 (Objects), (E) Experiment 5 (Objects – Short encoding condition), and (F) Experiment 5 (Objects – Long encoding condition). Bars with asterisks show comparisons between T1 and T2, and T2 and T3 that are significant; ** $p < .01$, *** $p < .001$.

ceiling at T1/T2 (no additional participants were available to re-run the T2 vs T3 contrast). This decreased the exclusion rate to 39 % for the T1 vs T2 contrast. The same pattern of results was seen, with no significant difference in dependency between T1 and T2 ($BF_{01} = 5.95$, $t_{69} = 0.72$, $p = .475$, $d = 0.09$).

5.3. Discussion

Experiment 4 replicated the results of Experiment 3 – retrieval dependency for object features did not decrease across delay. However, both Experiments 3 and 4 are at odds with the results of Experiment 2, where a decrease in dependency was seen across delay. Thus, the timing of the initial test phase (i.e., the change from immediate test to a 1 h delay for T1 between Experiments 2 and 3) is unlikely to have driven the differences between Experiment 2 and 3. Two further possibilities remain: (1) the encoding time of the events (shorter in Experiment 2 relative to 3 and 4) and (2) the delay of the final test session (3 days in Experiment 2 and 2 days in Experiments 3–4).

Due to time and resource constraints, we were unable to systematically test these potential discrepancies using the same event encoding paradigm as in Experiments 1–4. However, we were able to collect data using a similar paradigm, presenting only objects during encoding (no locations or people were presented). This approach is more similar to the original Brady et al. (2013) study, making the results of Experiment 5 potentially more comparable to their results. In this non-preregistered experiment, we varied object encoding time (1 s vs 3 s) and tested object memory at two delays (T1: immediate, T2: 5 h). Thus, we were able to assess the possibility that fragmentation of object features is more common for shorter than longer encoding times (one of the key differences between Experiments 2 and 3–4). We predicted that retrieval dependency should decrease across delay, however this decrease in dependency may be modulated by encoding time (i.e., a greater decrease should be seen in the 1 s vs 3 s encoding condition).

6. Experiment 5

Experiment 5 tested memory for object features immediately after encoding and after 5 h. Half of the objects were presented for 1 s at encoding and half for 3 s. We assessed retrieval accuracy and retrieval dependency as a function of both delay and encoding time. Note, in Experiment 2 participants had 5 s with the three elements presented, resulting in approximately 1.6 s per stimulus (presuming attention was evenly distributed across the three elements). They had a further 4 s of blank screen where they were instructed to imagine all three elements interacting in a meaningful manner. In Experiments 3–4, the stimuli were presented for 6 s, resulting in approximately 2 s per stimuli (with an additional 6 s of imagery). The 1 s encoding time in Experiment 5 is therefore shorter than any of the previous experiments, and the 3 s encoding time is longer than previous experiments. We adopted a Bayesian approach, collecting data until our contrasts of interest showed evidence (BF > 6) for or against the null hypothesis. Specifically, we were interested in whether retrieval dependency decreased across delay in both the short and long encoding condition, and whether this potential decrease in dependency was modulated by encoding time.

6.1. Method

6.1.1. Participants

Sixty-three participants were recruited for this two-part online experiment. Initially, the participants were recruited via a convenience sample for an undergraduate project but additional participants were subsequently recruited via Prolific (www.prolific.co). To be eligible for recruitment, there were no criteria initially apart from being young adults (18–40 years old). The Prolific participants additionally had to be native English speakers who have lived in the UK for 5 or more years, with normal or corrected-to-normal vision and no colourblindness. Of the participants who started the experiment, 45 participants (Mean age = 24.7, SD = 5.45; 30 female, 14 male) completed both sessions, met the inclusion criteria and were retained in the analyses. We lost demographic data for one participant, therefore the demographic information reported previously refer to 44 participants. Although this study was not preregistered, participants were excluded based on the eligibility criteria that we have used before in Experiments 2–4 on visual objects, outlined in the preregistration for Experiment 4 (https://osf.io/32zwm). If participants retrieved the correct object with an accuracy at or below 30 % or at or above 95 % in either of the two testing sessions (T1, T2) they were excluded from further analyses. Participants were also excluded if there were 6 or more successive timeouts in their memory responses. Of the 18 participants (28 %) who were excluded from analyses, 4 did not return for the follow-up testing session, nine had retrieval accuracy <30 % at T2, 4 had retrieval accuracy <30 % at T1, and 1 had too many timeout responses. Of the 45 remaining participants, there was a balanced number of participants in each of the counterbalanced study lists. We continued to recruit participants until we reached BFs > 6 for or against the null in relation to the difference in dependency between T1 and T2 in both the 1 s and 3 s encoding conditions (i.e., two separate Bayesian *t*-tests).

6.1.2. Design

Experiment 5 had a within-subjects design with the factor Delay (two levels: Test phase T1 and T2) and Encoding time (1 s and 3 s). The dependent measures of interest were retrieval accuracy and retrieval dependency for object features.

6.1.3. Procedure

This experiment used the same object stimuli as Experiments 1–4. The experiment consisted of a study session comprising an object encoding task with two different encoding blocks (i.e., a long block with trials lasting 3 s, and a short block with trials lasting 1 s; order counterbalanced across participants), a test session immediately after encoding (T1), and another test session after 5 h (T2).

In the study phase, participants saw 60 objects (30 per encoding block) presented one-by-one on a blank screen. On each trial, participants first saw a fixation point on the screen for 500 ms, followed by one image of an object for 1 s or 3 s, depending on the block. Participants were instructed to “study each object carefully” while it remained on the screen. There was a blank inter-trial interval of 500 ms before the onset of the next trial. Participants were given an opportunity to take a break in between the blocks. Note, Experiment 5 did not include the “imagine” period that was included in Experiments 1–4 so no imagery was explicitly required at encoding.

In the test phase, participants' memory for the object features was tested on separate blocks. In contrast to Experiments 1–4, object memory was tested without person/location cues (as no people/locations were presented at encoding) and participants were asked to select the encoded object in the correct state and colour. The target was presented among three foils from the same category with combinations of correct and incorrect features. Half of the short and long encoding time objects were tested at T1 (15 objects per condition, 30 trials in total, intermixed), and the other half at T2 (30 trials). On each trial, participants saw a fixation point on the screen for 500 ms, followed by the four response option images appearing on the screen. One of the response options was the target object, and three were foils of the same object in different combinations of state and colour. The order of the response options on the screen from left to right was random. The options remained on the screen for 6 s or until participants made a response by pressing one of the four keys (1–4). After a 500 ms blank interval participants saw a prompt “How confident are you in your choice?” on the screen, and again were asked to press the keys 1–4 to rate their confidence. The confidence rating prompt remained on the screen for 6 s or until response. There was a blank inter-trial interval of 500 ms before the onset of the next trial. There were several versions of the experiment based on the objects randomly selected for presentation either in the short or the long blocks of encoding trials, based on whether the short or long blocks appeared first during the study session, and based on which half of the objects appearing in short or long encoding blocks were tested immediately after encoding or after 5 h. All the different versions of the experiment were counterbalanced.

6.1.4. Analyses

6.1.4.1. Retrieval accuracy and confidence. Mean retrieval accuracy and mean retrieval confidence were calculated for the test phase at T1 and T2 for the short and long encoding time objects. Timeouts on memory trials were converted to incorrect responses for the purposes of calculating retrieval accuracy and dependency. 5.26 % (SD = 6.72 %) of the responses were timeouts at T1 and 2.96 % (SD = 4.94 %) were timeouts at T2.

6.1.4.2. Retrieval dependency. Retrieval dependency was calculated in the same manner as Experiment 4 for T1 and T2 and short and long encoding times separately, with the exception being that we only constructed one contingency table per condition (as each object was only tested on a single trial without person/location cues).

6.2. Results

6.2.1. Non-preregistered analyses

Participants remembered the object features at T1 for short (Mean accuracy = 0.58, SD = 0.19) and long (Mean accuracy = 0.67, SD = 0.16) encoding times and at T2 for short (Mean accuracy = 0.45, SD = 0.20) and long (Mean accuracy = 0.58, SD = 0.15) encoding times. Retrieval accuracy was above chance at T1 for short ($BF_{10} = 2.36 \times 10^{12}$, $t_{44} = 11.84$, $p < .001$, $d = 1.77$) and long ($BF_{10} = 4.53 \times 10^{18}$, $t_{44} = 17.84$, $p < .001$, $d = 2.66$) encoding times and at T2 for short ($BF_{10} = 499{,}953$, $t_{44} = 6.75$, $p < .001$, $d = 1.01$) and long ($BF_{10} = 5.76 \times 10^{15}$, $t_{44} = 14.88$, $p < .001$, $d = 2.22$) encoding times. There was evidence for a reduction in retrieval accuracy from T1 to T2 in both the short ($BF_{10} = 89.93$, $t_{44} = 3.94$, $p < .001$, $d = 0.59$) and long ($BF_{10} = 183.84$, $t_{44} = 4.19$, $p < .001$, $d = 0.63$) encoding conditions.

Evidence for retrieval dependency was seen at T1 for short ($BF_{10} = 7{,}332$, $t_{44} = 5.42$, $p < .001$, $d = 0.81$) and long ($BF_{10} = 5{,}976$, $t_{44} = 5.35$, $p < .001$, $d = 0.80$) encoding times and at T2 for long ($BF_{10} = 635$, $t_{44} = 4.62$, $p < .001$, $d = 0.69$) but not short ($BF_{10} = 0.28$, $t_{44} = 1.10$, $p = .279$, $d = 0.16$) encoding times. Comparing dependency between T1 and T2 we saw evidence for a significant decrease in dependency in the short encoding time condition ($BF_{10} = 10.75$, $t_{44} = 3.13$, $p = .003$, $d = 0.47$) but evidence for no decrease in the long encoding time condition ($BF_{10} = 0.16$, $t_{44} = 0.10$, $p = .925$, $d = 0.01$). Further, there was support in favour of a greater decrease in dependency over delay in the short relative to long encoding time condition ($BF_{10} = 2.71$, $t_{44} = 2.52$, $p = .015$, $d = 0.38$).

6.3. Discussion

Experiment 5 provided evidence for a decrease in retrieval dependency across delay when objects were encoded in isolation (i.e., without a simultaneously presented location and person), but only when the object was presented for 1 s. No decrease in dependency was seen across delay when objects were presented for 3 s. Thus, Experiment 5 provides evidence that encoding time might be a critical factor in determining whether object features are forgotten in a more fragmented rather than holistic manner.

7. General discussion

Across five experiments, we assessed whether memory for events and objects are holistically forgotten or fragment over time. Participants encoded 'events' by viewing images of three elements (i.e., a location, a famous person, and an object) and vividly imagining the three elements interacting. We measured retrieval accuracy and dependency for events and their constituent elements (locations, people, and objects) and objects and their features (state and colour) immediately and across different delays (from 1 h to 3 days). Despite clear evidence for forgetting of event elements over time, as measured by retrieval accuracy, we saw no evidence for a decrease in retrieval dependency (Experiments 1 and 3). This supports the proposal that events are forgotten in a holistic 'all-or-none' manner. The forgetting of object features was more variable. We saw evidence for a decrease in retrieval dependency in Experiments 2 and 5, however no decrease was seen in Experiments 3 and 4 (despite clear evidence of forgetting). Thus, we provide partial support for our prediction that object features fragment as a function of forgetting. We first discuss the holistic forgetting of events before focussing on the possible boundary conditions for the fragmentation of objects and their features.

7.1. The holistic forgetting of events

In Experiments 1 and 3, we saw clear evidence of forgetting but no evidence for a decrease in retrieval dependency when retrieving the elements of an event. Retrieval dependency is present when the retrieval of one element (e.g., person) is related to the retrieval of another element from the same event (e.g., object) when cued by the third event element (e.g., location). The presence of a statistical relationship between the retrieval of within-event elements has been taken as evidence for relatively coherent event representations that are retrieved in an 'all-or-none' manner (Horner & Burgess, 2013, 2014; Meiser & Bröder, 2002; Meiser, Sattler, & Weisser, 2008), though see (Starns & Hicks, 2005, 2008). Joensen et al. (2020) reasoned that if retrieval dependency didn't decrease over time, despite a clear decrease in retrieval accuracy, this could be taken as evidence that events were forgotten in a relatively holistic manner. At the extreme, those events that were forgotten were forgotten in their entirety and those that were remembered were remembered in their entirety. Thus, the statistical relationship between within-event elements remained even in the presence of forgetting. Here we switched from using verbal (written words) to visual images (i.e., of locations, famous people, and objects). Despite this switch in stimulus format, we replicated the results of Joensen et al., providing further support for the proposal that events are forgotten in a holistic manner (also see, Parra et al., 2024).

Across experiments, our 'events' consisted of three images of a location, famous person, and object. While most real-world events are likely built from these prototypical elements, real-world events are clearly more complex in nature. They are more multimodal, have actions and causal links, and can extend for longer periods of time. It is not currently clear whether our consistent findings of holistic forgetting apply to more complex events where the elements might be less well integrated. One possibility is that our 'events' represent what might be considered as the 'core' of an event representation – the key constituent elements – with other non-critical elements being more loosely associated. If true, it might be possible that fragmented forgetting can be seen for more non-critical elements that are none-the-less part of the event representation. Future research carefully constructing event representations with both critical and non-critical elements and measuring accuracy and dependency across these elements could assess this possibility.

The present experiments do not provide further insight into the mechanism that drives this holistic forgetting. We have proposed that the presence of coherent 'event engrams' in the hippocampus, retrieved via pattern completion, drives the retrieval dependency seen both immediately after encoding and after a delay (Horner & Burgess, 2014; Joensen et al., 2020). Forgetting, potentially due to decay (Hardt et al., 2013; Sadeh et al., 2014), may cause individual event engrams to become inaccessible, resulting in a failure to retrieve the entire event. This retrieval failure for the whole event would manifest as a decrease in retrieval accuracy but no decrease in retrieval dependency, the behavioural pattern we see in the present experiments. Further research is needed to test the prediction that decay, rather than interference, is driving the forgetting of events. The subsequent encoding of events with overlapping elements (e.g., the same location and person, but not object; similar to (Zotow, Bisby, & Burgess, 2020)) could test for possible interference processes, whereas decay would be expected to occur regardless of the subsequent encoding of overlapping events (Sadeh et al., 2016).

Given systems consolidation theories that predict decreased hippocampal involvement with increased delay (Alvarez & Squire, 1994; Born & Wilhelm, 2012; Dudai, Karni, & Born, 2015; McClelland, McNaughton, & O'Reilly, 1995), though see (Barry & Maguire, 2019; Nadel & Moscovitch, 1997; Winocur & Moscovitch, 2011; Yonelinas, Ranganath, Ekstrom, & Wiltgen, 2019), a further question is whether the retrieval dependency we see following a delay is driven by hippocampal pattern completion. We have recently shown that the hippocampus continues to contribute to the holistic reinstatement of event elements in the neocortex, alongside an emergent non-hippocampal contribution, at least after a 24-h period (Joensen, Ashton, Berens, Gaskell, & Horner, 2024). Thus, holistic retrieval is maintained despite the emergence of a non-hippocampal retrieval mechanism. Regardless of the underlying mechanism, the presence of holistic forgetting of events appears robust to the type of stimulus used at encoding and retrieval, the encoding time, and the delay. This lack of decrease in dependency for events across experimental factors is a critical finding when compared with the decreases in dependency for object features seen under some experimental conditions.

7.2. The (sometimes) fragmentary forgetting of objects

In Experiments 2–4 we embedded the encoding of objects in a wider 'event' that included a famous person and object. We then assessed retrieval accuracy and dependency for the features of the object (state and colour) immediately and following a delay. Whereas retrieval accuracy decreased across delay in all three experiments, retrieval dependency decreased in Experiment 2 but remained constant in Experiments 3–4. In Experiment 5, we presented objects in isolation during encoding and varied encoding time. We saw decreases in retrieval dependency across delay in the short encoding condition, but not the long encoding condition. Thus, we found evidence for both holistic and fragmented forgetting across Experiments 2–5.

Although the present studies are not able to reveal the precise boundary conditions for object fragmentation (see possible explanations below), the finding that objects do fragment under certain conditions is critical. First, it contrasts with the event memory findings, where evidence for fragmentation is not seen under a variety of experimental conditions (Experiments 1 & 3 in the current manuscript; Joensen et al., 2020; Parra et al., 2024). The finding that objects can fragment, but events cannot, points to a dissociation where object representations are more susceptible to fragmentation than event representations. Second, the object memory literature currently has seemingly contradictory findings with evidence for both fragmented (Brady et al., 2013) and holistic forgetting (Balaban et al., 2020; Parra et al., 2024) of object features. Our results help to reconcile these findings, suggesting that they are both correct – objects can be forgotten in a more holistic or fragmented manner depending on the experimental conditions. Further, Experiment 5 points to one specific boundary condition for fragmentation – encoding time. There are likely to be multiple interacting factors that cause fragmentation. Providing a full understanding of these boundary conditions would require a wider programmatic series of studies.

One possible explanation for the differences between objects and events is that evidence for object fragmentation are false positives, with more studies showing holistic than fragmentary forgetting. This seems unlikely, given evidence for fragmentary forgetting in two out of the four experiments that tested memory for object features in the present studies, and the evidence for fragmentation in Brady et al. (2013). Further, in Experiment 2 the Bayes Factor was 8.64 in favour of a decrease in retrieval dependency across delay and in Experiment 5 for the short encoding time it was 11.58. Thus, it seems more likely that fragmented forgetting of object features does occur, but only under specific experimental conditions. A key question for future research therefore is what are the boundary conditions that allow for fragmented (or holistic) forgetting of object features?

The first possible boundary condition is the delay between encoding and retrieval. The delay in Experiment 2 was three days, relative to two days in Experiments 3–4. However, both Brady et al. (2013) and Balaban et al. (2020) used a three-day delay and found conflicting evidence for fragmentation. Further, in Experiment 5 we found evidence for fragmentation with a 5 h delay. Although we cannot rule out a role for delay in relation to fragmentation, it seems unlikely that this factor contributed to the differences between experiments (at least in a simple direct way that doesn't interact with other factors).

The second possible boundary condition is encoding time. Experiment 2 presented all three images for 5 s, followed by 4 s blank screen where participants were required to "imagine the elements interacting". This contrasts with Experiments 3 and 4 that presented images for 6 s followed by a 6 s imagination period. Thus, participants had less time to encode and imagine the object (and person/location) in Experiment 2, where fragmentation was seen, relative to Experiments 3–4, where fragmentation did not occur. Similarly, in Experiment 5, objects (in isolation) were presented for either 1 s or 3 s, with fragmentation only occurring in the short encoding condition. If attention was split evenly between the three elements over time in Experiments 2–4, participants had approximately 1.7 s to encode the object in Experiment 2 relative to 2 s in Experiments 3–4. It is possible that the 1 s (Experiment 5) and 1.7 s (Experiment 2) encoding time led to object features being forgotten in a more fragmented manner relative to 2 s (Experiments 3–4) and 3 s (Experiment 5) encoding. However, the difference in encoding time between Experiments 2 and 3–4 is relatively short (~300 msec per element, 1 s total encoding time for all three elements). Further research would need to systematically vary the encoding time of events (similar to Experiment 5, but over a wider range of times) to directly test this proposal.

If encoding time is a critical factor, we do not believe it is simply that less encoding time results in poorer encoding, leading to fragmentation. Instead, less encoding time is likely to lead to a qualitatively different, perhaps less coherent or integrated, object representation that is more susceptible to fragmentation. First, retrieval accuracy for objects at T1 in Experiment 2 (0.59) was higher than for objects in Experiment 3 (0.55), and evidence for fragmentation was seen in the former but not the latter (though note that T1 in Experiment 3 included a short 1 h delay). Thus, lower initial retrieval accuracy, which presumably would suggest poorer encoding, does not appear to increase the likelihood of fragmentation. Second, we performed an analysis where we correlated retrieval accuracy at T1 with the difference in dependency between T1 and T2 for object features across participants. We found no evidence for a relationship. Thus, it seems unlikely that fragmentation is driven by something as simple as poorer encoding.

The third possibility is that encoding objects embedded within more complex 'events' increased the extent to which object features were bound together, decreasing the likelihood of fragmentation over time. This possibility might have been further increased by the presence of an "imagination" period during each encoding trial following stimulus presentation. The presence of multiple event elements, and the active imagination of these elements interacting, may have increased the likelihood of the object features being included in a higher-order event representation. Given evidence that event representations are forgotten in a more holistic manner, if object features were bound within this higher-order representation this could explain the presence of holistic forgetting of object features (in Experiments 3–4). This proposal fits with recent evidence that attending to the thematic 'story' between sequentially presented visual images (relative to attending to stylistic details) boosts retention for both semantic and perceptual details of the images (Vijayarajah, McAlister, & Schlichting, 2022). It also fits with the lack of decrease in dependency for object features seen in Parra et al. (2024), given the objects tested were embedded in real-world novels that included higher-order event and narrative content. The presence of fragmentation in Experiment 2 does not fit with this account, although it is possible that the shorter encoding and imagination time in Experiment 2 meant this binding was less likely to occur. Further, this proposal does not account for the presence of holistic forgetting in Experiment 5 (in the long encoding time condition) and in Balaban et al. (2020), where objects were encoded in isolation.

It has been proposed that object representations, supported by the neocortex, are more susceptible to interference than event representations (Sadeh et al., 2014), and that feature specific interference might be a mechanism that induces object fragmentation (Andermane et al., 2021). Further studies that systematically induce interference for specific object features, for example by encoding new objects with similar colours to previously encoded objects, are needed to directly test this prediction. Coupled with the experiments proposed above where participants encode events with overlapping elements, these studies would directly test the degree to which interference induces both forgetting and fragmentation across objects and events.

Interestingly, our exploratory analyses in Experiment 3 showed higher levels of retrieval dependency for objects relative to events. How initial levels of dependency relate to decreases in dependency over time is another unexplored question. Regardless of the precise experimental conditions for object fragmentation in Experiments 2 and 5 and holistic forgetting in Experiments 3–5, a further question relates to how objects and their features are represented. There is evidence in the literature for the independent representation of object features (Oberauer & Eichenberger, 2013; Olson & Jiang, 2002; Reinitz et al., 1992) as well as object features being more holistically bound (Ceraso et al., 1998; Duncan, 1984; Vogel et al., 2001; Wilton, 1989). More recently, it has been proposed that objects can be represented in either manner, dependent on the encoding conditions (Kuhbandner, 2020). Further, it has been proposed that objects are represented hierarchically, where object features might be represented independently lower in the hierarchy but bound together higher in the hierarchy (Brady et al., 2011; Li et al., 2022). This hierarchical proposal for objects fits with the broader literature, including our own proposal related to forgetting of episodic memories (Andermane et al., 2021), that episodic events (which include objects and object features) are represented in a hierarchical manner (Robin & Moscovitch, 2017; Sekeres, Winocur, & Moscovitch, 2018). The precise conditions at encoding, including (but not limited to) encoding time and whether objects are encoded during a mental imagery period along with other event elements, will drive the likely fragmentation of object features. The presence of any form of higher-order binding, whether it is the binding of features in a single object representation in the perirhinal cortex (Barense et al., 2007; Bussey & Saksida, 2007; Li et al., 2022) or the binding of the object and its features within a higher-order event representation in the hippocampus (Andermane et al., 2021; Horner & Doeller, 2017), will likely drive a more holistic form of forgetting. When objects are encoded in isolation, with little imagery, for shorter periods of time, then such higher-order representations are perhaps less likely to form, resulting in more fragmentary forgetting.

One final factor across experiments was a high exclusion rate. Although we cannot rule out that excluding a high proportion of participants biased our results in some manner, this possibility seems unlikely. First, irrespective of exclusions, we saw consistent levels of dependency across delay for events and these results replicate those from Joensen et al. (2020). For objects, we saw evidence for both fragmented and holistic forgetting and critically this variability is not readily explained by differences in exclusions. Although Experiments 3 and 4 had the highest exclusions rates and showed holistic forgetting for objects, our exploratory analyses where we included more participants showed the same pattern of results. Finally, it is important to note that we exclude on the basis of retrieval accuracy but our main theoretical conclusions relate to differences in retrieval dependency. If participants are not at ceiling or floor in relation to accuracy (our exclusion criteria assured this was the case) then the dependency measure is not related to accuracy (as it explicitly controls for differences in accuracy at the individual participant level).

7.3. Conclusion

Across five experiments, we have assessed the forgetting of events and objects using a single experimental approach. We have shown that events and their elements are forgotten holistically, whereas objects and their features can sometimes be forgotten in a fragmentary manner. Our results point to a dissociation between objects and events where object representations are more susceptible to fragmentation than event representations. Further, evidence for both more holistic and fragmentary forgetting of object features helps to reconcile contradictory findings in the object memory literature. The precise boundary conditions that result in object fragmentation are still unclear and will require a programmatic body of research to fully understand, however we have proposed that certain encoding conditions (in particular, encoding time and mental imagery) may encourage the binding of object features into more coherent object and/or event representations that drive more holistic forgetting. Our theoretical account and current experiments bring together the fields of object memory and episodic memory, providing evidence for a hierarchical framework of episodic memory and forgetting that includes object representations and allows for both fragmentation and holistic forgetting of object features. Critically, our results point to a dissociation between the forgetting of events and objects – forgetting of events is consistently holistic whereas forgetting of objects is sometimes fragmentary.

CRediT authorship contribution statement

Nora Andermane: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arianna Moccia:** Writing – review & editing, Visualization, Validation, Software, Formal analysis, Data curation, Conceptualization. **Chong Zhai:** Methodology, Investigation, Formal analysis. **Lisa M. Henderson:** Writing – review & editing, Supervision, Conceptualization. **Aidan J. Horner:** Writing – review & editing, Visualization, Supervision, Software, Methodology, Funding acquisition, Conceptualization.

Acknowledgements

This research was funded by an Economic and Social Research Council (ESRC) grant award to AJH and LH (ES/R007454/1). AJH is also funded by the ESRC (ES/X00791X/1) and Leverhulme Trust (RPG-2023-003). For the purposes of open access, the authors have applied a creative comments attribution (CC-BY) license to any author accepted manuscripts version arising henceforth. All preregistrations, data, and statistical analyses are available on the Open Science Framework (https://osf.io/zgra9/).

Data availability

Data and analysis code is fully shared and available on OSF and links are provided in the manuscript

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T., van Langen, J., & Kievet, R. A. (2021). Raincloud plots: a multiplatform tool for robust data visualization. *Wellcome Open Research, 4*, 63.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the Natlional Academy of Sciences, 91*, 7041–7045.

Andermane, N., Joensen, B. H., & Horner, A. J. (2021). Forgetting across a hierarchy of episodic representations. *Current Opinion in Neurobiology, 67*, 50–57.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science, 319*(5870), 1640–1642.

Balaban, H., Assaf, D., Arad Meir, M., & Luria, R. (2020). Different features of real-world objects are represented in a dependent manner in long-term memory. *Journal of Experimental Psychology: General, 149*, 1275–1293.

Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia, 45*(13), 2963–2974.

Barry, D. N., & Maguire, E. A. (2019). Remote memory and the Hippocampus: A constructive critique. *Trends in Cognitive Sciences, 23*(2), 1–15.

Berens, S. C., Richards, B. A., & Horner, A. J. (2020). Dissociating memory accessibility and precision in forgetting. *Nature Human Behaviour, 4*, 866–877.

Berron, D., Schutze, H., Maass, A., Cardenas-blanco, A., Kuijf, H. J., Kumaran, D., & Duzel, E. (2016). Strong evidence for pattern separation in human dentate gyrus. *Journal of Neuroscience, 36*(29), 7569–7579.

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General, 147*(2), 243–256.

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: A new self-report imagery questionnaire. *Applied Cognitive Psychology, 20*(2), 239–263.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76*(2), 192–203.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 1–34.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General, 142*(3), 791–808.

Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus, 17*(9), 898–908.

Ceraso, J., Kourtzi, Z., & Ray, S. (1998). The integration of object properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1152–1161.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, Amnesia and the hippocampal system*. MIT Press.

Cooper, R. A., & Ritchey, M. (2019). Cortico-hippocampal network connections support the multidimensional quality of episodic memory. *eLife, 8*, Article e45591.

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron, 88*(1), 20–32.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501–517.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*.

Fisher, J. S., & Radvansky, G. A. (2019). Linear forgetting. *Journal of Memory and Language, 108*.

Grande, X., Berron, D., Horner, A. J., Bisby, J. A., Düzel, E., & Burgess, N. (2019). Holistic recollection via pattern completion involves hippocampal subfield CA3. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 39*(41), 8100–8111.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10−11), 1409–1422.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111–120.

Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J. W., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications, 6*(7462), 1–11.

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General, 142*(4), 1370–1383.

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology, 24*(9), 988–992.

Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology, 43*, 102–109.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. *The American Journal of Psychology, 35*(4), 605–612.

Joensen, B. H., Ashton, J. E., Berens, S. C., Gaskell, M. G., & Horner, A. J. (2024). An enduring role for hippocampal pattern completion in addition to an emergent nonhippocampal contribution to holistic episodic retrieval after a 24 h delay. *Journal of Neuroscience, 44*(18).

Joensen, B. H., Gaskell, M. G., & Horner, A. J. (2020). United we fall: All-or-none forgetting of complex episodic events. *Journal of Experimental Psychology: General, 149*(2), 230–248.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21*(11), 1551–1556.

Kuhbandner, C. (2020). Real-world objects are represented in visual long-term memory both as unbound features and as bound objects. *Frontiers in Psychology, 11*, 3454.

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience, 33*(13), 5466–5474.

Li, A. Y., Fukuda, K., & Barense, M. D. (2022). Independent features form integrated objects: Using a novel shape-color “conjunction task” to reconstruct memory resolution for multiple object features simultaneously. *Cognition, 223*, Article 105024.

Marks, D. F. (1973). Visual imagery differences in the recall of picture. *British Journal of Psychology, 64*(1), 17–24.

Marr, D. (1971). Simple memory: A theory for Archicortex. *Philosophical Transactions of the Royal Society of London B, 262*(841), 23–81.

Martin, C. B., Douglas, D., Newsome, R. N., Man, L. L., & Barense, M. D. (2018). Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. *eLife, 7*, Article e31873.

Mayes, A. R., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences, 11*(3), 126–135.

McClelland, J. L., McNaughton, B. L., & O’Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419–457.

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39*(4), 352–370.

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 116–137.

Meiser, T., Sattler, C., & Weisser, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 32–49.

Motley, S. E., & Kirwan, C. B. (2012). A parametric investigation of pattern separation processes in the medial temporal lobe. *Journal of Neuroscience, 32*(38), 13076–13084.

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus’ forgetting curve. *PLoS One, 10*(7), Article e0120644.

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology, 7*(2), 217–227.

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*(2), 416–427.

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science, 30*(12), 1696–1706.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition, 41*(8), 1212–1227.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the ‘strong-object’ hypothesis. *Perception & Psychophysics, 64*(7), 1055–1067.

Parra, D., Antes, N., & Radvansky, G. A. (2024). Event cognition and holistic versus fragmented remembering and forgetting. *Proceedings of the Annual Meeting of the Cognitive Science Society, 46*.

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition, 20*(1), 1–11.

Robin, J., & Moscovitch, M. (2017). Details, gist and schema: Hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences, 17*, 114–123.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734–760.

Sacripante, R., Logie, R. H., Baddeley, A., & Della Sala, S. (2022). Forgetting rates of gist and peripheral episodic details in prose recall. *Memory & Cognition, 51*, 71–86.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*(1), 26–36.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2016). Forgetting patterns differentiate between two forms of memory representation. *Psychological Science, 27*(6), 810–820.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry, 20*(11), 11–21.

Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters, 680*, 39–53.

Spachtholz, P., & Kuhbandner, C. (2017). Visual long-term memory is not unitary: Flexible storage of visual information as features or objects as a function of affect. *Cognitive, Affective, & Behavioral Neuroscience, 17*(6), 1141–1150.

Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1213–1220.

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review, 15*(2), 309–314.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioural Neuroscience, 100*, 147–152.

Tulving, E. (1983). *Elements of episodic memory*. Clarendon Press.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*(1), 49–60.

Utochkin, I. S., & Brady, T. F. (2020). Independent storage of different features of real-world objects in long-term memory. *Journal of Experimental Psychology: General, 149*(3), 530–549.

Vijayarajah, S., McAlister, E., & Schlichting, M. L. (2022). Encoding-phase orientation toward thematic content over perceptual style benefits picture memory. *Memory*, 1–11.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114.
Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*(2), 225–252.
Wilton, R. N. (1989). The structure of memory: Evidence concerning the recall of surface and background colour of shapes. *The Quarterly Journal of Experimental Psychology Section A, 41*(3), 579–598.
Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society, 17*(5), 766–780.
Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*(10), 515–525.
Yonelinas, A. P., Ranganath, C., Ekstrom, A. D., & Wiltgen, B. J. (2019). A contextual binding theory of episodic memory: Systems consolidation reconsidered. *Nature Reviews Neuroscience, 20*, 364–375.
Zotow, E., Bisby, J. A., & Burgess, N. (2020). Behavioral evidence for pattern separation in human episodic memory. *Learning & Memory, 27*(8), 301–309.