Cognition 266 (2026) 106317

Contents lists available at ScienceDirect

Cognition

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

Full Length Article

Causal coherence improves episodic memory of dynamic events

Andreas Arslan *, Jonathan F. Kominsky

Central European University, Department of Cognitive Science, 51-55 Quellenstraße, 1100 Vienna, Austria

ARTICLE INFO

Keywords:
Episodic memory
Causal cognition
Event representation
Event perception

ABSTRACT

"Episodes" in memory are formed by the experience of dynamic events that unfold over time. However, just because a series of events unfold sequentially does not mean that they are related. Sequences can have a high degree of causal coherence, each event connecting to the next through a cause-and-effect relationship, or be a fragmented series of unrelated occurrences. Are causally coherent events remembered better? And if coherence leads to better recall, which attributes of episodic memories are particularly affected by it? Past work has investigated similar questions by manipulating the causal structure of language-based, narrative stimuli. In this study, across three experiments, we used dynamic visual stimuli showing unfamiliar events to test the effect of causal structure on episodic recall in a cued memory task. Experiment 1 found that the order of three-part causally coherent sequences of events is better remembered than that of fragmented events. Experiment 2 extended this finding to longer sequences and further demonstrated that causal structure is not confounded with low-level characteristics of the stimuli: Reversing the order of coherent stimuli led to task performances indistinguishable from those on fragmented stimuli. Experiment 3 replicated the results of improved order recall from the previous experiments and additionally showed that recall of causally relevant details of coherent stimuli is superior to recall for details of focal events in fragmented sequences. In sum, these findings show that the episodic memory system is sensitive to the causal structure of events and suggest coherence usually leads to better recall.

1. Introduction

Memories of past experiences are usually incomplete and sometimes distorted. Even the most vivid fail to recapture past situations perfectly: We neither remember every detail originally perceived in a given moment, nor the exact temporal arrangement of all such moments taken together. While a certain loss of information appears inevitable as immediate experiences are transformed into memories, it is evident that not all experiences are remembered equally well.

Compare, for example, two types of event: Standing in a kitchen and watching a friend prepare a dish, versus sitting in a public park, looking at the people walking past you. If we contrast how the different parts making up these experiences relate to each other across time, we find that the two events are quite dissimilar in their degree of *causal structure*. One consists of a sequence of actions and outcomes that can be described in terms of causes and effects, i.e., it is causally *coherent*. The other is made up of largely independent observations whose temporal arrangement does not affect their interpretation, i.e., it is causally *fragmented*. There seems to be a clear difference between events that are almost automatically thought of as a succession of causes and effects (e.g., tapping an egg firmly against the rim of a container will lead to it cracking, which allows the egg to be whisked) and those we feel just happen to occur in a certain order (e.g., someone pushing a stroller through the park, and *then* a police officer rides by on horseback).

Here, we investigate whether the causal structure of events influences their representation in memory and how accurately they can be recalled. Specifically, we study the impact of causal structure on episodic memories, which are characterized as those memories that represent past situations in a relatively holistic format that captures multiple aspects of the original experience (Tulving, 1972, 1983). We focus on this subdivision of the declarative memory system (Squire, 2004), because we are interested in how the identification of causal relations *during* the experience of an event might shape its representation in memory.

Causal structure provides relations between different events that may support memory for causally relevant content, and for the order in which events occur. Turning back to our original comparison of cooking vs people-watching, say our friend breaks a few eggs, pours them into a

* Corresponding author at: Department of Cognitive Science, Central European University, Room D510, 5th Floor, 51-55 Quellenstraße, 1100 Vienna, Austria.
E-mail addresses: Arslan_Andreas@phd.ceu.edu (A. Arslan), kominskyj@ceu.edu (J.F. Kominsky).

https://doi.org/10.1016/j.cognition.2025.106317
Received 2 April 2025; Received in revised form 31 August 2025; Accepted 1 September 2025
Available online 11 September 2025

bowl, whisks the mass, and then transfers it into a hot pan. While these individual events can all be thought of in isolation, when observed in this order they are obviously not independent from each other. It then may be the case that remembering one part reinforces the memory of others that are linked to it. For instance, if you correctly remember eggs being cracked into a pan, by basic causal and logical necessity you would have no difficulty remembering that the eggs being removed from the carton came before this event, and the scrambling of the eggs in the pan happening afterward. If, on the other hand, the causal structure of an event is low, recalling its content may provide hardly any information about the order of events. You could remember everyone you saw in the park perfectly, but even though this would give you a comprehensive list of impressions (jogger, person pushing a stroller, police officer) it contains no clues as to the order in which they passed by. If your memory for the order of events is enhanced by causality, it also implies that you might have enhanced memory for causally relevant *content* of those events (so that you can order them). For example, if you remember the scrambled eggs in the pan, it may help you also remember the whole eggs that were cracked into the pan, or at least stop you from misremembering the whole eggs as a saltshaker (see Altmann & Ekves, 2019 for a similar analysis, albeit in a different theoretical context). By contrast, when sitting on a park bench and seeing a jogger run by, and then someone pushing a stroller, and then a police officer on horseback, none of these individual observations has any bearing on the other, they are only associated by coincidence, so remembering the participants in one observation would not help you remember anything about the participants of the others.

Previous research on the possible influence of causal structure on memory seems to have been largely confined to how it affects the comprehension and recall of (mostly written) narratives. Bartlett (1932) showed that ghost stories that defy rational interpretation are often misremembered, Bransford and Johnson (1972) found that a lack of context makes narratives both difficult to understand and recall, and multiple other studies manipulating causal relations on the word or sentence level suggested that they affect how well a text is recalled (Black & Bern, 1981; O'Brien & Myers, 1987; Radvansky & Copeland, 2000; Radvansky, Tamplin, Armendarez, & Thompson, 2014; Trabasso & Van Den Broek, 1985). Going beyond textual narratives, Poulsen, Kintsch, Kintsch, and Premack (1979) had 4- and 6-year-old children recreate picture stories that were presented in either a normal or scrambled arrangement. Children in both age groups showed better retention of the correctly arranged stories that featured clear causal relations. Taken together, this prior work shows that causal information affects how conventional narratives are interpreted and remembered.

Nearly all of these previous studies, however, relied on written texts or static images, meaning the causal structure and content of events were conveyed to participants in a rather abstract manner different from first-hand perceptual experience. For instance, take a sentence like 'A person runs up to a ball and kicks it.': While the kind of event the sentence *refers to* is certainly dynamic, it is underspecified[1] in many features that would be effortlessly extracted from a directly observed dynamic event (e.g., whether they paused before kicking, how fast or far they ran, etc.). Whether the causal structure of dynamic events that are *perceived* in real-time influences their recall has, to our knowledge, not yet been systematically investigated.

Another reason previous findings may not generalize to dynamic events is that they rely on stimuli that follow conventionalized narrative structures. Although narrative and discrete temporal structure often go hand-in-hand (e.g., in cartoons and novels), stimuli that allow for the temporally continuous depictions of events and are at face value more 'naturalistic' (e.g., live-action videos) might still feature familiar narrative tropes and techniques. When drawing on professionally made movies or movie clips to study the relation between the causal structure of events and memory (see, for instance, Brownstein & Read, 2007; Lee & Chen, 2022; Antony, Lozano, Dhoat, Chen, & Bennion, 2024), it is fundamentally very difficult to distinguish the influence of on-line causal inferences on recall from individuals' ability to rely on their familiarity with highly conventionalized narrative schemas. While narrative structure is an important dimension of many events and previous studies therefore deliberately made it an essential part of their stimuli, such approaches leave ambiguous whether memory benefits from causal structure per se or, or only in conjunction with familiar narrative schemas.

1.1. Inferring causal structure

In the context of this study, ascribing a 'high degree of causal structure' to a dynamic event is shorthand for 'most individuals experiencing the event will interpret it as a sequence of causally related parts.' In other words, the degree of causal structure indicates how much causally relevant information is available to an individual during an experience and whether that information can likely be organized into a coherent representation. Individual causal relations between events can be identified on-line either through low-level perceptual mechanisms (Michotte, 1963) or rapid inference. Previous studies already showed that rapidly inferred causal relations frequently have postdictive effects that result in *false* memory. When watching clips of action events like kicking a ball, participants tended to falsely report having seen moments of physical contact (e.g., the tip of the shoe touching the ball) even when they had not actually been shown, but only if there was a 'causal implication' (Strickland & Keil, 2011). In the example of kicking a ball, this causal implication consisted of showing its continued trajectory. If participants instead saw something like a person walking away, they were less likely to erroneously 'fill in' the moment of contact. Importantly, a subsequent study showed that this effect is not due to the influence of familiar, stereotyped event schemas (like kicking a ball) as it appears in unfamiliar, novel events (Kominsky, Baker, Keil, & Strickland, 2021).

The fact that this works with unfamiliar and novel events illustrates how causal structure is not exactly the same as predictability. Events can have high predictability in the absence of direct causality (e.g., a meal's main course is followed by a dessert) or exhibit clear causal structure but be effectively unpredictable (e.g., the outcome of a coinflip). While past studies have explored the relationship between predictability and memory for dynamic events (Exton-McGuinness, Lee, & Reichelt, 2015; Sinclair & Barense, 2018; Sinclair, Manalili, Brunec, Adcock, & Barense, 2021; Wahlheim, Eisenberg, Stawarczyk, & Zacks, 2022), they did not explicitly manipulate or separately explore the role of causal structure.

1.2. The causal backbone of memory episodes – Coherent and fragmented events

At its most basic, an episodic memory is simply a representation of an event that preserves some its experiential dimensions (Tulving, 1972, 1983) rather than completely 'abstracting them away,' as a semantic memory representation might. Because there is no consensus on what does or does not, cognitively speaking, constitute an 'event' (Yates, Sherman, & Yousif, 2023), there is no a priori lower bound on the degree of causal structure any one experience might exhibit: Many situations may really be like the people-watching example in that they largely consist of a stream of impressions whose order is more or less arbitrary. Yet it is also inarguable that many situations are not, and that everything from seeing a window being slammed shut by a gust of wind to watching a multi-step experiment in chemistry class can be justifiably called a 'dynamic event with a high degree of causal structure'. Here, we hypothesize that inferring and representing these causal dependencies should allow for more accurate recall of these events.

[1] A recent study by Bigelow, McCoy, and Ullman (2023) suggests people tend to oftentimes 'leave out' rather basic features of scenes when asked to imagine them vividly.

Causal structure, whenever it can be inferred and remembered, necessarily imposes a temporal direction on a recalled situation that moves from causes to effects. Moreover, it by definition guarantees enough detail is retained to identify the approximate nature of the causal relation. Encoding the causal structure of an event could act as a means of compressing representations in episodic memory. Take the memory of how you watched how your friend

1) firmly tapped an egg against the rim of a porcelain bowl,
2) which led to a crack forming in the shell,
3) through which yolk started to flow out.

Each event is in some way dependent on its immediate predecessor, but any change to their order would disrupt those dependencies. This is a hallmark of what we call causally coherent events: They exhibit causal structure that is lost if the order of event parts or the participants of those events are altered. Of course, not all objects and object features are causally relevant. While confusing the egg with a saltshaker would contradict the overall causal relations in the above example, falsely remembering it was a brown instead of a white egg would not.

Tracking the causal relations between parts of a coherent event could more efficiently accomplish what would otherwise require explicit enumeration of contiguous parts: the creation of a single structured representation that preserves the order and relevant details of a past event. As mentioned before, however, many sequences of events need not be coherent. When sitting in the park and looking at the stream of passersby, what is perceived is not the result of readily inferable causal relations – neither with respect to order nor content. We call these events causally fragmented. Because they lack causal structure that might organize the resulting episodic memory representation, we hypothesize that people are less likely to remember their precise order and content, compared to events with a more coherent causal structure.

1.3. The present experiments – Unfamiliar events with varying causal structure

As of yet, there is no direct empirical evidence (to our knowledge) showing that the causal structure of dynamic events affects their representation in episodic memory. While previous findings in the discrete and explicit realm of conventional language-based narratives point in this direction, they may not translate to events that unfold in continuous time and arrive in the form of 'unlabeled' sensory input. Moreover, those prior studies tended to not distinguish between predictability and causal structure which, as argued above, often co-occur but are nonetheless separable. It is therefore unknown whether event sequences that encourage the real-time inference of causal connections between their components are remembered differently from sequences of apparently unrelated events.

To investigate whether this is the case and if it improves the accuracy of episodic memories of such events, we employed visual stimuli that are devoid of conventional narrative structure. Moreover, to ensure that any potential effect found would be as content-independent as possible, we created videos of unfamiliar scenarios that were unlikely to call any specific learned event schemas to mind (e.g., cooking a meal, or playing baseball). We intended to create conditions in which participants instead would have to rely on their basic knowledge of physical principles governing most real-life situations (Spelke, Breinlinger, Macomber, & Jacobson, 1992), to perceptually extract causal information from motion and collision events. Studies of people's intuitive understanding of physics have shown that adults have expectations about the strength of gravity and conservation of energy following collisions (Twardy & Bingham, 2002), expect visual entropy ('disorder') to be indicative of how much time has elapsed (Clarke & Tyler, 2024), and, relatedly, are able to detect when an anomalous event appears to run counter 'the arrow of time,' especially when critical cues are present (Hanyu, Watanabe, & Kitazawa, 2023). This small selection of findings illustrates that humans are equipped with intuitions[2] about motion and time that should enable them to identify simple cause and effect relations in entirely unfamiliar situations, as long as they abide by known laws of (macroscopic) physics.

In all three experiments reported here, the stimuli presented to participants were either causally *coherent* or causally *fragmented*. These stimuli were videos that featured one or more objects in motion and consisted of multiple short clips separated by cuts, but they differed with regard to what happened during the transitions between clips. In coherent videos only the viewer's perspective changed between cuts; the 'global state' of the event was carried over from one clip to the next (e.g., rather than watching a friend cooking in the room, this resembled watching a cooking show on TV that switches between multiple camera angles). In contrast, in fragmented videos the consequences of events shown in one clip were 'reset' in the next, in that object transformations were undone or spatial relations altered in a way that was incompatible with the previous clip (Fig. 1). Fragmented videos therefore included causal relations *within* each clip but, critically, not *between* clips (analogous to the people-watching example). The main goal was not to contrast causal and acausal events, but examine the effect of causal coherence on the retention of multi-part events. Participants' episodic memory was tested in a cued memory task.

Experiment 1 sought to investigate whether there are systematic differences in recall between causally coherent and fragmented videos that were made up of three clips. We tested both how well order was recalled and how well participants could distinguish non-episode from episode content (whether a still image belonged to the same video as a cue image). Experiment 2 intended to replicate the findings of Experi-

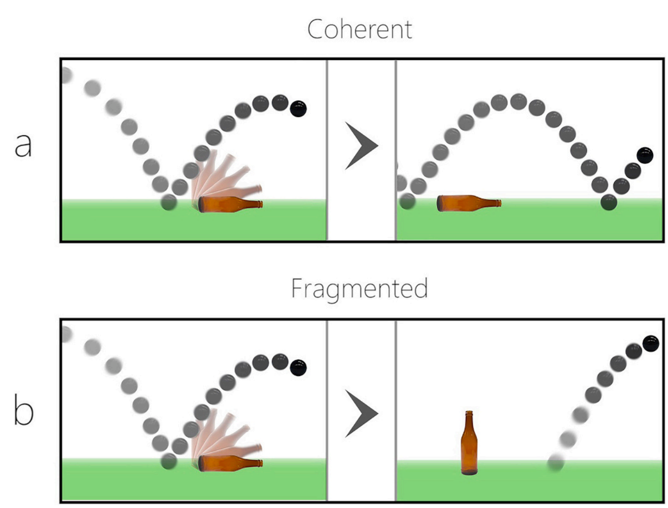

Fig. 1. All stimuli consisted of multiple consecutively shown events. **a** Coherent stimuli were those where whatever happened in one event (the bottle falling over after having been hit by a ball) carries over to the subsequent event (the bottle lying on the ground and the ball continuing to bounce). **b** Fragmented stimuli featured 'resets' or abrupt changes of positions, velocities etc. whenever one event transitions to the next. Here, the fallen bottle is upright again after the transition, and the ball's direction of motion is different from what it was previously.

[2] These capacities have been proposed to, in their totality, constitute something akin to a mental 'physics engine' and it has been suggested that certain state changes in that internal, 'video game'-like physics simulation could mark causal relations, such as collisions (Ullman, Spelke, Battaglia, & Tenenbaum, 2017). Other studies, however, indicate that humans' simulative capacities are in some respects quite limited (Ludwin-Peery, Bramley, Davis, & Gureckis, 2020; Ludwin-Peery, Bramley, Davis, & Gureckis, 2021).

ment 1 and to examine whether those findings hold for longer events and across events of variable length (operationalized as the number of clips in each video). In fact, we hypothesized that longer coherent events may be remembered *better* than short coherent ones because they would contain a greater number of sequentially dependent causal relations. The supposition was that if causal structure has an effect that can be described as something like a compression of memory representations, it should be detectable for longer events as it, presumably, would be of greatest functional relevance in such contexts.

In Experiment 3 we investigated whether coherent causal structure affects how well causally relevant *content* is preserved in episodic memories. We hypothesized that not everything would be remembered more accurately but that there would be better memory for *causally relevant* details. Experiment 3, in addition to testing memory of event order, probed participants' ability to reject doctored still images (lures) that deviated from parts of videos they had actually seen. Lures manipulated only the position and/or presence of objects central to the events shown in the video stimuli. Because these objects could only have an overarching causal significance to multiple parts of a video in the coherent condition, we predicted that participants would also be better at identifying and rejecting them in that condition.

2. Experiment 1

Experiment 1 sought to investigate whether events that are interpretable as sequences of causes and effects are remembered better overall. To test this hypothesis, we employed a cued memory task. All of the stimuli were motion events in 3D Euclidian space, comprising three short consecutive clips. The three clips in a causally coherent video stimulus showed a complex event evolving forward in time from multiple viewing angles, but in the causally fragmented stimuli the final state (position, velocity, etc.) in one clip did not carry over to the next. To give an example (illustrated in the upper right quadrant of Fig. 2): Clip *one* of a fragmented video shows a green cube (and others not shown) falling into a black cup, with an L-shaped object visible in the background. In clip *two*, the previously inert L-shaped object hovers in midair and knocks over the black cup – when it subsequently rolls across the floor, only one yellow cube is revealed to be inside. These spatiotemporal disruptions are methodologically inspired by Kominsky et al. (2021). Critically, both coherent and fragmented videos had spatiotemporally disruptive cuts that changed the viewing position of the objects, and only differed in whether the causal consequences of the clip before the cut carried over into the clip following the cut or were 'reset.' This was a deliberate choice to control for the role of spatiotemporal disruptions; otherwise, the causally coherent videos would be very similar in their low-level visual features from one clip to the next, while the fragmented videos would be very different.

Besides investigating whether causally coherent events are remembered better overall, Experiment 1 allowed us to differentiate two measures indicative of memory structure: The ability to accurately recount the order in which events occurred, and the ability to correctly identify whether something (here: an image) was part of a particular cued event or not. In our task, participants first saw a cue image from one clip in a video, followed by a number of test images. They then had to judge whether each test image was part of the cued episode (an 'in-episode' item), and if so whether it came before or after the cue image, or if the image was unrelated to the cued episode (a 'non-episode' item). Assuming that causal structure helps with the individuation of episodes,

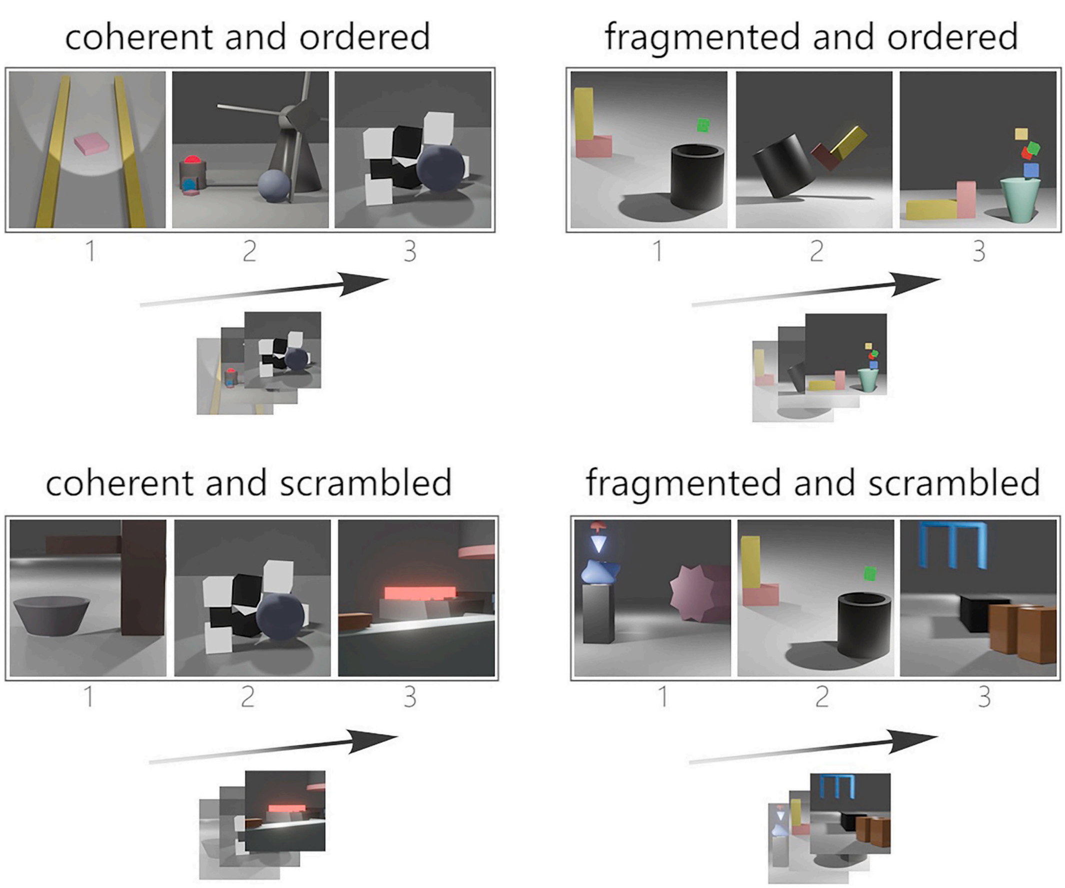

Fig. 2. Each video consisted of three short clips, which are represented as still images in this figure. Shown in this static manner, the differences between these stimulus conditions are not readily apparent, but it may be noted that the scrambled videos in the bottom row include a greater variety of objects, as they are assembled from the clips of three different videos. Stimuli for Experiment 1 can be viewed at: https://osf.io/r5p27/?view_only=4ac5cd76d99b4c82b5f20bd0f41d02cd

we hypothesized an interaction between the effects of *causal status* (coherent vs. fragmented) and *episode status* (in-episode vs. non-episode) on memory accuracy.

2.1. Method

Experiment 1 and the two experiments following it were in accordance with the principles of the Declaration of Helsinki and approved by Central European University's Psychological Research Ethics Board (PREBO). Informed consent was obtained from all individual participants included in the study. No identifying information is included in either the manuscript itself, supplemental materials, or the data and materials made available in online repositories.

2.1.1. Participants

Sixty adult participants took part in Experiment 1 (Mean age = 27.4; 29 female, 31 male). To calibrate task difficulty, we conducted two pilot experiments with a sample size of 10 each. Based on the observed effect size in the second pilot comparing overall accuracy between the coherent and fragmented conditions (Cohen's $d = 0.43$), we conducted a power analysis that indicated that 60 participants would be sufficient to reach >90 % power to detect this effect. Our only exclusion criterion was a completely uniform response profile (pressing exclusively one of the three possible response keys across all 64 test items) and as no participant exhibited such a decision pattern all of them were included in the analysis.

All participants were recruited via Prolific (https://www.prolific.com) and received £3 for their participation. We did not specifically target any group or demographic; fluency in the English language and no previous participation in pilot studies associated with the experiment were our only prerequisites. Prospective participants gave informed consent via a Qualtrics survey (https://www.qualtrics.com) and then completed the actual experiment on the Pavlovia platform (Bridges, Pitiot, MacAskill, & Peirce, 2020; https://www.pavlovia.org). The median completion time for the experiment was 22.09 min.

2.1.2. Materials and procedure

We created 16 video stimuli in the 3D graphics software Blender (v3.5.1 and immediate predecessors, The Blender Foundation, www.blender.org, 2023). Half of these videos were designed to be causally *coherent*, the other half causally *fragmented*. All videos were composed of three clips/segments of equal duration, with the total duration of each video amounting to 11.37 s. Clips were demarcated by a cut that invariably entailed an instantaneous, discontinuous displacement of the virtual camera. While we expected that the cuts and three-clip structure would be noticeable to participants in the majority of movies, this may not have always been the case. The task in the subsequent test phase did not require participants to be cognizant of the videos' partitioning.

Both coherent and fragmented videos largely depicted motion and collision events involving a variety of objects. The animation of these events relied on Blender's inbuilt tools for physical simulation (mainly 'Rigid Body World'). Objects featured in the videos ranged from realistic and detailed models (tire, figurine, spoon) to simple three-dimensional geometric shapes (spheres, cubes, prisms). Most videos followed the movements of a focal object across multiple clips. Although these focal objects were frequently familiar items, they were not placed in contexts that conformed to potential schemas strongly associated with them, but rather arbitrarily juxtaposed. Only relatively small sections of the virtual environment were shown at a time as the camera usually stayed close to whatever happened to be the focal object at the moment. This ensured that the *overall* progression of events depicted in the videos was unpredictable: Objects ended up interacting with other objects or elements of the environment that hitherto had remained unseen (see for instance Fig. 2, top left).

All videos were purposely set in a grayish space without any recognizable features or landmarks in the distant background. Due to the aforementioned fact that much of the motion in the videos was simulated rather than keyframed, the virtual camera frequently had to move quickly and along complicated trajectories to keep the action in frame. Because this was the case for coherent and fragmented videos alike, three is no reason to assume that this occasional failure to follow cinematographic conventions like the 180-degree-rule should skew experimental results one way or another. At most, the hectic nature of the videos may make it harder to infer the relevant causal connections in coherent stimuli.

Crucially, coherent and fragmented videos did not contrast on the level of individual clips but the manner in which they related to each other. Changes in the position, velocity, shape or of any other physical attribute of an object in one clip of a coherent video always carried over to the subsequent clip (Fig. 1a). By contrast, in fragmented videos, the configurations of the objects in the scene were never fully preserved across clips (Fig. 1b). While the same shapes and focal objects recurred in each of a video's three clips, each cut brought with it a discontinuous rearrangement of these elements and, occasionally, even sudden state changes, such as the reconstitution of an object that shattered or was deformed in the previous clip. Both coherent and fragmented videos exhibited a wealth of causal dependencies in individual clips. Only coherent videos, however, depicted events that spanned all three clips and therefore enabled an observer to infer a causal structure that connects even the non-adjacent first and final clip.

To rule out that the transitions between the three clips of the videos were not more 'jarring' in the fragmented condition, we performed a simple analysis of the pixel-level difference between the final frame before a cut and the first frame following it on all stimuli of both groups. We did this by subtracting the frame after the cut from the one preceding the cut using functions available in the OpenCV library (Bradski, 2000) and then summed up the values of the resultant image array to obtain a scalar. After calculating a mean pixel-level transition difference for each video, we performed an independent samples *t*-test and found no significant difference between coherent and fragmented stimuli, $t(14) = 0.353$, $p = 0.729$.

Besides causal status, we manipulated the arrangement of video's constituent clips, creating movies that were either *ordered* or *scrambled*. Ordered videos were precisely those described above, whereas scrambled ones were created from clips of three different unrelated videos (Fig. 2, bottom row). Not only were scrambled videos composites of multiple videos, the position of a clip (it being the first, second or third clip) in a scrambled video need not have corresponded to the position it had in the original video it was extracted from. It is worth noting that the descriptions 'coherent and scrambled' and 'fragmented and scrambled' simply indicate what type of clip was used to create scrambled composites: either coherent or fragmented ones (never both). Because they were assembled in a purposely arbitrary way, scrambled videos were no longer 'coherent' in the above sense of changes carrying over from one clip into the next. Likewise, scrambled fragmented videos were necessarily more erratic than their ordered counterparts, as now not only the state and configuration of objects changed discontinuously between cuts, but entirely different objects were the focus of attention in each clip (compare the top right image and bottom right image in Fig. 2).

Participants watched 16 videos in total, eight of which were coherent and another eight fragmented. Furthermore, half of the 16 videos were ordered and the other half scrambled. Distributing these properties symmetrically resulted in four videos that were coherent and ordered, four that were fragmented and ordered, four that were coherent and scrambled, and four that were fragmented and scrambled.

Pilot experiments made it clear that presenting all videos sequentially in one block prior to the test trials would lead to chance-level performance in the memory task. We therefore divided the experiment into four blocks, with each one consisting of four video presentations followed by a test phase. Every block featured videos from each of the four conditions (coherent/ordered; coherent/scrambled; fragmented/ordered; fragmented/scrambled) in random order. To rule out confounding effects arising from the idiosyncratic details of specific videos, all participants were randomly assigned to one of two stimulus groups. These stimulus groups differed only in regard to which videos were scrambled and which were ordered. The videos that participants in one stimulus group saw as scrambled composites were seen in their 'original,' ordered state by those in the other, and vice versa. Videos were randomly assigned to one of the four blocks ahead of the experiment, meaning the set of videos shown in each block was identical for all participants within a given stimulus group. The order of presentation within a block was randomized for each participant.

Ahead of the first presentation phase, participants were simply told that they were about to see videos and instructed to watch them attentively. The four videos displayed in each block were separated by a two-second gap, during which participants saw a numerical countdown onscreen. Following this presentation, participants underwent a cued memory task. The cue was a frame (still image) taken from one of the videos they had watched during the presentation phase. 50 % of these cue images were extracted from the second clip of a video, 25 % from the first and another 25 % from the third clip. The cue was paired with the instruction to remember where they had seen this image. They could look at it for as long as they wished before continuing. Participants then saw four test items, which were presented to them consecutively, in random order. Each test item was a still image like the cue. Participants had to press one of three relevant keys on the keyboard of their device to report whether they thought an image belonged to a part of the video preceding the cue ('b' key), a part of the video that came after the cue ('a' key), or a video unrelated to the cue ('x' key).

The task phase in each block required participants to make 16 such decisions (four test items for each of the four cue images), amounting to a total of 64 trials for the entire experiment and making 64 the maximum score a participant could reach. As we varied the *episode status* of the test items, every set of four test items associated with a cue included two non-episode images that were unrelated to the video the cue originated from. This means 'x' was the correct choice 50 % of the time. The remaining in-episode test items were equally split regarding whether 'before' or 'after' was the correct answer (i.e., 25 % of all answers apiece).

Between the videos and the test images in the first block there was a brief practice session after participants had watched the four videos. This practice session served the purpose of familiarizing participants with the task and the relevant keys. They first saw two extremely simple three-part videos, each of which consisted of static geometric shapes, like a triangle, sphere or cube. Afterwards, they could practice the cued memory task and (in contrast to the actual task phases) received feedback on each choice, informing them whether it had been correct or not, and, in the latter case, specifying why it had been wrong (i.e., "this was not in the same video" or "this was presented after the cue image, not before").

2.2. Results and discussion

The initial dependent variable in all subsequently described analyses was the participants' score on the cued memory task, i.e., the sum of correct choices in the cued memory task, ranging from 0 to 64. Because some statistical tests were conducted on only a relevant subset of questions, we report *accuracies*, which are just the proportion of correct responses, to ease comparability. Our first preregistered (https://osf.io/g7xuv/?view_only=61cd830e0429456c9b33de4af29f398f) analysis consisted in comparing the average performances on test items with different *causal status* (coherent vs. fragmented). To this end, we conducted a paired-samples *t*-test using the Python library *scipy* (Virtanen et al., 2020). The 64 test items were evenly split in terms of causal status, with 32 cued by frames from coherent and the remaining 32 cued by frames from fragmented videos. The mean proportion accuracy for coherent items ($M = 0.60$, $SD = 0.16$) was significantly higher than that for fragmented items ($M = 0.53$, $SD = 0.14$), $t(59) = 4.93$, $p < 0.001$, $d = 0.42$.

In another preregistered analysis, we examined the overall impact of *arrangement* (ordered vs. scrambled) on accuracy in the same fashion. The 64 test items were again equally divided into ordered and scrambled ones. A paired samples *t*-test indicated that the mean accuracy for ordered items ($M = 0.67$, $SD = 0.18$) significantly exceeded that achieved on scrambled items ($M = 0.46$, $SD = 0.14$), $t(59) = 12.88$, $p < 0.001$, $d = 1.28$.

Our final preregistered analysis investigated the effect of both causal status and *episode status* (in-episode vs. non-episode) of items on accuracy. 'In-episode' items were those images in the cued memory task that were stills from the same video (the same 'episode') as the cue itself, 'non-episode' items were images belonging to a different video than the cue. For all non-episode items the correct response was pressing the x-key, for the in-episode items it was either the a- or b-key. We conducted a 2 (causal status) x 2 (episode status) repeated measures analysis of variance (ANOVA) which collapsed across the factor of arrangement. We observed no significant interaction between the two factors, $F(1, 59) = 0.98$, $p = 0.32$, but each factor on its own significantly affected accuracy: It was higher for coherent items than for fragmented ones, $F(1,59) = 24.32$, $p < 0.001$, $\eta_p^2 = 0.29$ and higher for non-episode items than in-episode items, $F(1,59) = 7.29$, $p = 0.009$, $\eta_p^2 = 0.11$.

We conducted a 2 (causal status) x 2 (episode status) x 2 (arrangement) repeated measures ANOVA using R's afex package (Singmann et al., 2019), which we had preregistered as an exploratory analysis (see Supplementary table 9 for the full array of results). This revealed a three-way interaction, $F(1,59) = 37.71$, $p < 0.001$, $\eta_p^2 = 0.39$. Because performance in the scrambled condition was of no relevance to our overarching hypotheses beyond what we had already found, we subsequently conducted a 2 (causal status) x 2 (episode status) ANOVA on *only ordered items*, using the Python library *Pingouin* (Vallat, 2018). Here, episode status, $F(1,59) = 20.8$, $p < 0.001$, $\eta_p^2 = 0.26$ and causal status, $F(1,59) = 9.28$, $p = 0.003$, $\eta_p^2 = 0.13$ continued to have a significant effect on accuracy, and furthermore we found a significant interaction between the two factors, $F(1,59) = 13.24$, $p < 0.001$, $\eta_p^2 = 0.18$. As the left panel in Fig. 3 illustrates, accuracy for non-episode items is the same regardless of causal status, whereas accuracy for in-episode items is higher for items that are also coherent than for those that are fragmented.

These results indicate that participants remembered the order of clips in coherent events more accurately. Because this was a novel task design, we wanted to rule out that something besides memory could account for the performance difference. Specifically, we wanted to address the concern that the coherent still images shown in the cued memory task could potentially provide more information about order than the fragmented images. If this were the case, the viewing of the videos and participants' memory of them would be irrelevant, and reconstruction of the event based on still images alone could drive the performance difference between the groups. We therefore ran another version of Experiment 1 (preregistered here: https://osf.io/mutfc/?view_only=685c68dac99b4e7594d320a499352d75) that did not feature a presentation phase; i.e., no videos were shown to the participants. In every other respect (besides instructions that had to be reworded) the experiment was identical to what was described in the method section. We tested 59 participants. The decisive finding of that follow-experiment was that accuracy on coherent ordered in-episode items was now at chance level ($M = 0.33$, $SD = 0.18$; compared to $M = 0.66$ and $SD = 0.23$ in the main experiment), showing clearly that the stills alone do not enable participants to reconstruct the order of coherent events (see SI for further analyses and discussion of 'Experiment 1.1').

In summary, in Experiment 1 participants showed better memory for coherent stimuli than fragmented stimuli, and this effect was driven by better performance on ordered in-episode items (i.e., remembering the order in which clips were presented). In accordance with our prediction, video stimuli with a causal structure that manifests as an unbroken chain

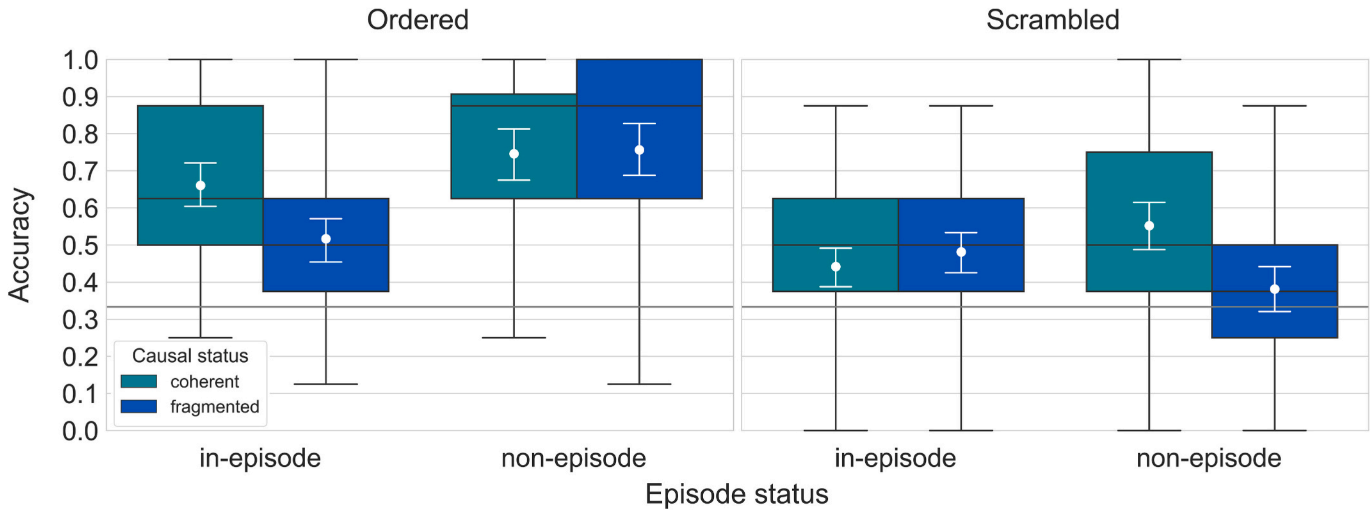

Fig. 3. Results of Experiment 1. Boxplots of accuracy across conditions. The left panel shows the results most relevant to our hypotheses: Accuracy was higher for coherent items. Furthermore, an interaction with episode status can be observed: Accuracy for coherent items was only higher if the episode status of an item was *in-episode*. Boxes indicate quartiles, whiskers the spread of the data (1.5 x interquartile range), and the black line median accuracy. White dots represent mean accuracy and the error bars around them 95 % confidence intervals (CIs). The horizontal gray line marks presumed chance-level performance (0.33). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of causes and effects spanning all three clips are better recalled than stimuli whose causal structure is fragmented. We also found that stimuli created by combining three clips from three different videos with no regard for order or causal structure – 'scrambled' stimuli – are associated with worse performance than ordered stimuli overall, suggesting a disruptive effect on memory.

The comparable performance on coherent and fragmented ordered *non*-episode items indicates that causal structure does not help individuals distinguish which element belongs to a particular memory and which does not. The superior performance on coherent ordered *in*-episode items, however, suggests that causal coherence leads to a more explicit encoding of event order. While this was something of a surprise to us (see hypothesis 2 in the preregistration), there is a plausible explanation for these findings: To correctly report that a still image does not originate from the same video as the cue image it is not necessary to represent the sequence of events depicted in that video. It may even be sufficient to rely on visual similarity – a strategy that is equally applicable to coherent and fragmented stimuli and test items. Determining whether an item came before or after the cue, however, is obviously not possible without recollecting the order of events to at least some extent. Thus, causal coherence primarily affected before/after judgments, but did not influence participants' ability to identify whether a clip was from the same video or a different video as the cue.

With Experiment 1 having established that causal structure measurably affects recall, we decided to follow it up with two experiments that each focused on different but related aspects of the phenomenon in question: In Experiment 2, we sought to clarify and extend our finding regarding the effect of coherence on the recall of event order. Experiment 3, on the other hand, focused on the potential influence of coherence on the retention of causally relevant content (in addition to order).

3. Experiment 2

The previous experiment showed that recall of events with coherent causal structure is in some respects superior to that of those with fragmented structure. Nonetheless, it is important to acknowledge that the videos we presented to participants in Experiment 1 were short in duration and all of equal length (three clips).

Experiment 2 then was designed to test whether the effect of better recall of causally coherent stimuli registered in Experiment 1 would be even more pronounced for longer sequences. Previously, participants' memory of videos comprising three clips was probed with regard to whether they could distinguish non-episode from in-episode imagery and recall the order of events. Three was chosen because it is the lowest number of subdivisions giving rise to a central clip which stands in both a 'before' and 'after' relation to other parts of the video. Another reason for the low number of clips was the considerable task difficulty made apparent by the fact that participants in Experiment 1 were well below ceiling performance.

The fact that any memory system's capacity is limited and recall for three-clip videos was already limited may suggest that the increased informational load inevitably tied to a larger number of clips in a video should push participants to perform at chance level. If, however, the episodic memory system is generally marked by a tendency to preserve the causal structure of events in the representations it generates, it may not be appropriate to think of additional clips as a linearly (or, with respect to explicit order-relationships between clips, potentially exponentially) increasing burden. It is possible that the more extended causal structure in a multi-part event allows for a more parsimonious or compressed encoding of an experience.

To investigate this, we presented videos with a variable number of clips to participants and tested their recollection of them with a cued memory task very similar to the one used in Experiment 1. Of main interest was whether causal status (*coherent* or *fragmented*) and number of clips (three, five or seven) showed an interaction. If causal structure is conducive to forming more extensive, temporally-indexed representations of events, performance should diverge as the number of clips increases, with coherent stimuli showing an increased advantage with more clips. Because the first experiment indicated that causal coherence affects memory of event order rather than the ability to distinguish memory from non-memory content, we simplified the task and only asked participants to make 'before vs after' judgments.

3.1. Method

3.1.1. Participants

As in Experiment 1, we aimed to recruit sixty adult participants (Mean age = 26.6; 34 female, 25 male) via Prolific. This sample size was simply chosen to match the previous number of participants. The requirements were fluency in English and no prior participation in pilot experiments or any other of our experiments similar to the present one.

We again would have excluded participants who responded with the same keypress (here: either 'a' or 'b') to every question from analysis but found that no participant gave completely uniform answers. Because one participant experienced technical problems while attempting to complete the online experiment and produced no salvageable dataset, our final sample size was $N = 59$. Participants received £3.75 and the median completion time was 23.45 min. As in the previous experiment, participants were asked for their consent in a Qualtrics survey, completed the behavioral task itself on the Pavlovia platform, and were finally debriefed (and informally asked about their impressions) in a second Qualtrics survey.

3.1.2. Materials and procedure

Experiment 2 inherited many of Experiment 1's characteristics; the stimuli used, however, were all newly created (again with the 3D-graphics software Blender). Instead of 16 videos with three clips each, the stimulus pool of this experiment comprised 12 seven-clip videos. The total duration of each complete video was 20.42 s, 2.92 s per clip.

Stimuli were again evenly split with regard to their causal status: six were causally coherent, the other six causally fragmented. Coherence and fragmentation were operationalized as described in Experiment 1. To allow for a more systematic creation of cue images, the event progression depicted in the videos was more constrained than in Experiment 1. Each video featured two *focal objects*, with focal object #1 being present in clips one through five and focal object #2 being present in clips three through seven. There were various other recurring objects or background features visible in each video, but only focal objects exhibited this pattern of appearance. Furthermore, a focal object was always in motion for at least some time when shown alone in a clip; in clips featuring both focal objects, at least one of them was in motion for at least part of the clip. As in Experiment 1, the videos showed unfamiliar situations that did not correspond to any identifiable narrative schema. To keep the overall predictability of coherent videos[3] low, the virtual environment of each video again was only gradually revealed as focal objects moved through it and interacted with it.

As mentioned, all of the 12 videos created for this experiment were designed to be seven clips long. However, not all participants saw the full-length version of each video, as the *number of clips* shown was an independent variable in Experiment 2. There were three different possible lengths: three, five, or seven (the 'entire' video). Videos with fewer than seven clips were generated from the complete original by applying a simple, symmetrical cropping rule. To obtain a five-clip video, the outermost two clips (the first and the seventh) were removed; the same held for a length of three, where the outermost four clips (first, second, sixth, and seventh) were removed, leaving just the three clips at the center (Fig. 4).

We further varied the *temporal* arrangement of stimuli by presenting half of them in order (*forward*) and half of them in reverse (*reversed*). 'Reversal' here does not signify an inversion of order on the level of individual frames within a clip but refers to the sequence in which the clips were presented within the video as a whole. In other words, clips could be presented in ascending numerical order (clip 1 then clip 2 then clip 3 etc.) or descending (clip 7 then clip 6 then clip 5 etc.), but the clips themselves were unchanged.

To ensure that cuts in certain kinds of videos were not considerably more jarring than in others, we conducted the same kind of pixel-level analysis of transitional frames described in the method section of Experiment 1. The results (including those of an identical analysis of the stimuli used in Experiment 3) are tabulated in section 4 of the SI; they give no indication of asymmetries in transition patterns likely to bias performance in one direction or another.

The experiment was divided into four blocks, each of which involved the presentation of three video stimuli and a task phase. Forward and reversed stimuli were presented in separate blocks. Every block included three videos, one of each length (3, 5, and 7 clips). Half of the blocks were composed of one fragmented and two coherent stimuli, the other half of one coherent and two fragmented stimuli. To ensure that all possible lengths of a video stimulus appeared in all possible states of temporal arrangement and causal status, we randomly assigned each participant to one of multiple stimulus groups, as in Experiment 1 (now six possible groups instead of two).

The cued memory task following each stimulus presentation was similar to that in Experiment 1, with some exceptions. First, the cue image here was always a still taken from the central (fourth) clip of a video stimulus. Second, the number of test items associated with each cue now depended on the number of clips of the corresponding video stimulus. If the video stimulus had a length of three, there were only two test items – one from the clip before and one from the clip after the middle clip. For videos with a length of seven, there were three items with 'before' as the correct answer, and three more where 'after' was correct. Participants could press *a* for 'after' and *b* for 'before' on their keyboards; there was no rejection option in this experiment.

3.2. Results and discussion

The dependent variable was again the *accuracy* achieved in the cued memory task, which could take on values between 0 and 1. To test the effect of causal coherence on memory, we first conducted a 2 (causal status: coherent vs fragmented) x 2 (temporal arrangement: forward vs reversed) x 3 (length/number of clips: 3, 5, 7) repeated measures ANOVA. All analyses reported in this section were preregistered (https://osf.io/2yt6k/?view_only=181b7e5b210d417b920e54dbdadbcbf4). The analysis showed no main effect of causal status itself, $F(1, 58) = 2.4$, $p = 0.11$, but a significant interaction between the factors causal status and temporal arrangement $F(1, 58) = 11.77$, $p = 0.001$, $\eta_p^2 = 0.17$, which was expected: temporal arrangement should not matter for causally fragmented videos (see Supplementary table 10 for the full array of results).

To better understand this two-way interaction, we next compared accuracies obtained on *forward* coherent items with that on *reversed* coherent items (ignoring number of clips). If reversing the sequence of clips shown in a video stimulus disrupts their causal structure, performance on reversed coherent items should be worse than on forward coherent items. A paired samples *t*-test showed the mean accuracy for forward coherent items ($M = 0.73$, $SD = 0.18$) was significantly higher than that for reversed coherent items ($M = 0.60$, $SD = 0.20$), $t(58) = 4.32$, $p < 0.0001$, $d = 0.68$. Critically, temporal arrangement had no detectable effect on fragmented items, with accuracies for forward fragmented items ($M = 0.61$, $SD = 0.17$) and reversed fragmented items ($M = 0.62$, $SD = 0.16$) not differing significantly, $t(58) = -0.33$, $p = 0.74$. It can be noted that the average accuracy for reversed coherent items was around the same value of that on both forward and reversed fragmented items.

To test whether number of clips had an effect on memory for causally coherent events, we subsequently conducted a 2 (causal status) x 3 (number of clips) repeated measures ANOVA, looking only at the performance on items with *forward* temporal arrangement (Fig. 5). (The contrast in the reversed condition is not of interest because causal status had no effect on memory for reversed items.) The dependent variable was again accuracy on this half of the items. We found a strong main effect of causal status, $F(1, 58) = 12.72$, $p < 0.001$, $\eta_p^2 = 0.18$, with coherent items leading to higher accuracy. Number of clips also significantly affected accuracy, $F(1,116) = 3.31$, $p = 0.04$, $\eta_p^2 = 0.05$. However, accuracy did not decrease as the number of clips increased (Fig. 5). Post hoc, uncorrected pairwise comparisons showed no significant difference between 3-clip and 5-clip videos, $t(58) = -1.98$, $p = 0.05$, and no significant difference between 5-clip and 7-clip videos, $t(58) = -0.11$, $p = 0.91$, but did show a significant difference between 3-clip

[3] Fragmented videos already are inherently unpredictable as every cut leads to arbitrary spatial rearrangements.

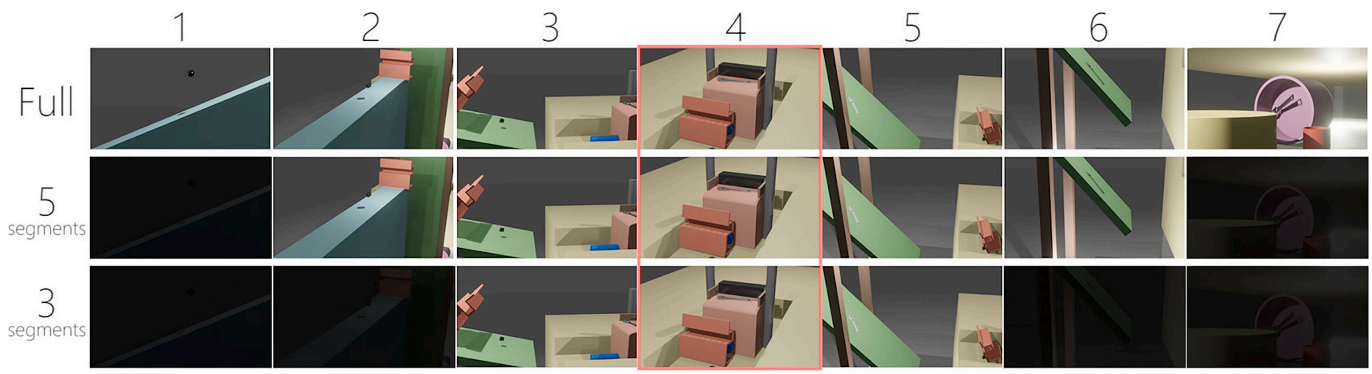

Fig. 4. The top row shows a full video, those below the cropped five and three clip long versions. Cue images were always sourced from the fourth and central clip (red rectangular border). Stimuli for Experiment 2 can be viewed at: https://osf.io/f47u8/?view_only=2c40f9ddc632478dab8f4dd01d7b9d77.

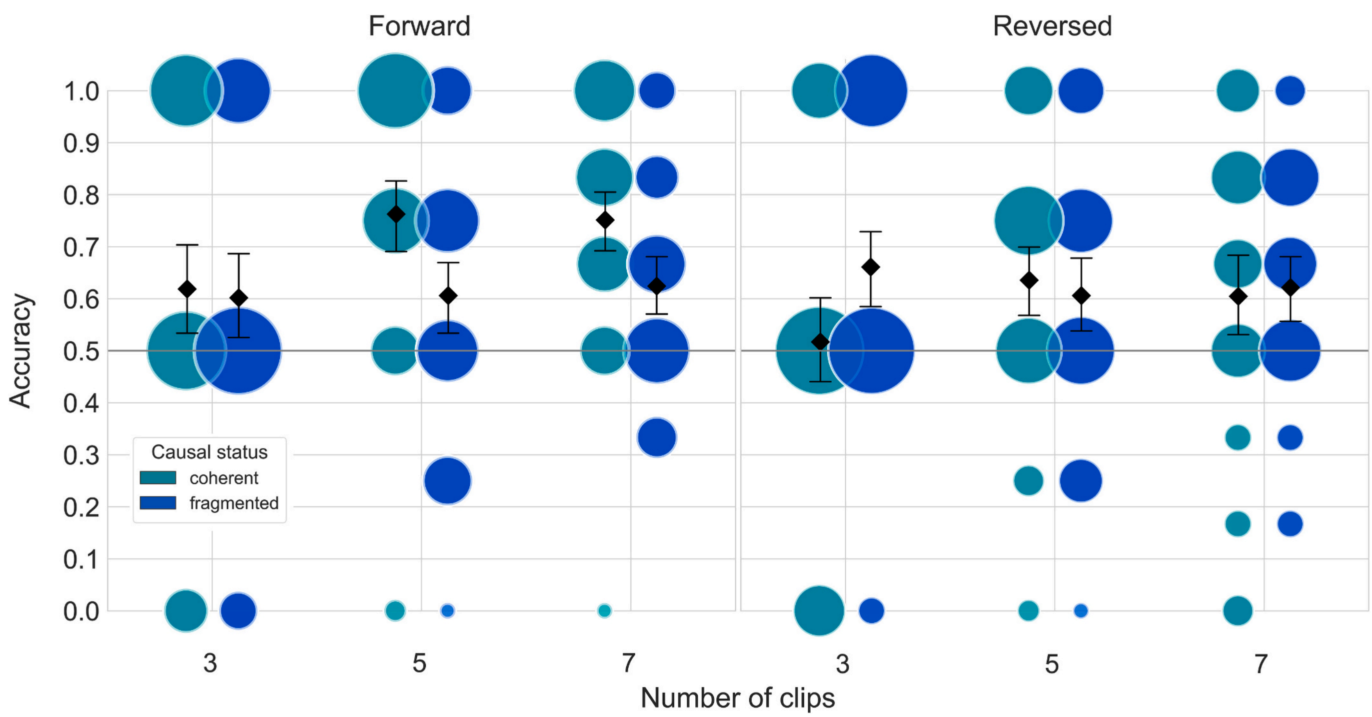

Fig. 5. Results of Experiment 2. Black diamonds represent mean accuracy; error bars represent 95 % CIs. A 2 (causal status) x 2 (temporal arrangement) x 3 (number of clips) repeated measures ANOVA revealed a significant interaction between causal status and temporal arrangement, with accuracy scores being highest on items that are *coherent* and *forward*. The left panel (forward temporal arrangement) shows that participants reached overall significantly higher accuracy on items with coherent causal status compared to those with fragmented causal status. Performance did not decline as number of clips increased. Colored disks represent individual participants' accuracies; their size indicates the relative frequency of a given value. Note that the number of test items was the number of clips - 1, so possible accuracy scores for 3-clip items were 0, 0.5, or 1.0 (0, 1, or 2 correct), whereas more intermediate accuracies were possible for 5- and 7-clip items. The horizontal gray line indicates presumed chance-level performance (0.5).

and 7-clip videos, $t(58) = -2.29$, $p = 0.03$, indicating an overall uptick in performance as number of clips increased. Contrary to our predictions, there was no significant interaction between causal status and number of clips, $F(1,116) = 2.62$, $p = 0.08$.

The overall pattern of results obtained in Experiment 2 is surprising in that it both runs counter to the particular assumptions that motivated it, but also offers more evidence for a positive effect of causal coherence on the representation of event order in episodic memories. We did not find an interaction between causal status and length that would suggest that causal coherence is disproportionately advantageous when participants have to remember longer causally connected events, at least with the range of lengths we tested. To conclusively test it, it would perhaps be necessary to use longer stimuli than we employed here. After all, participants performed well in general: They showed above-chance accuracy even for fragmented items across all levels of length. Without making the clips long enough to reduce performance to chance on fragmented items, it is more difficult to detect whether causal coherence would attenuate this decrease.

There was a slight surprising divergence from Experiment 1: When analyzed separately, there is no effect of causal coherence on three-clip videos with forward temporal arrangement, which is the most analogous condition to Experiment 1's stimuli. Although we have no definitive explanation for this result, at least three factors may have contributed to it: First, despite having the same number of clips, the shortest videos in Experiment 2 (8.76 s) lasted a few seconds less than those in Experiment 1 (11.37 s) and were therefore perhaps harder to follow or simply not as memorable. Second, and relatedly, in Experiment 1 all videos were of the same duration, whereas in Experiment 2 long and short videos were presented together within each block. Less attention may have been allocated to shorter videos in this changed context; maybe they even were explicitly judged to be less 'relevant' by participants. Third, in Experiment 2 the number of test items associated with each video was proportional to its number of clips: 6 test items for long videos, 4 for intermediate videos, and only 2 for short videos. The performance in the memory task is therefore less informative for short videos and errors are much more consequential.

As a whole, Experiment 2 replicates the finding that the order of subevents of causally coherent events is remembered more accurately than that of fragmented ones. Because we used an entirely new set of videos this suggests that our earlier findings are not attributable to particular features of the stimuli employed in the first study. Moreover, we could show that causal coherence does have an effect on the episodic memory of events that are substantially longer than those presented in Experiment 1.

4. Experiment 3

The purpose of Experiment 3 was to test whether causally relevant details of coherent events are better remembered than similar details of fragmented events. Experiment 1 already tested people's ability to reject images that were completely unrelated to a given episode (non-episode items). Correct responses there, however, were not informative as to what content was remembered with precision: Non-episode items showed multiple objects and environments that did not match the cued episode, hence recall of any one of these, or a general sense of a 'contextual mismatch,' could underlie correct rejections. Experiment 2 focused entirely on order, leaving Experiment 3 to address the open question of whether coherent causal structure enhances recall of specific (causally relevant) content.

We came to this hypothesis by considering an inherent asymmetry between recall of order and recall of content: While it is possible to accurately remember details of an event without recalling anything about its order (for example, one might say with confidence that a certain actor or famous building was or was not in a movie but have no idea about the sequence of events making up its plot), it is impossible to make correct order judgments without taking content into account. Put simply, content is what is being ordered, and must therefore take precedence. Of course, not all 'content' in a scene is informative with regard to order; *causally* relevant content, which is subject to the rule that causes must precede effects, is informative by definition. Thus, the superior recall of the event order of coherent videos we found in the first two experiments might be partly explained by superior recall of causally relevant details. In addition, recent work by Shin and Gerstenberg (2023) suggests that the causal relevance of a feature predicts whether it will be recalled and that participants' memory errors often consist in confusing a feature that was actually seen with one that is equivalent in terms of its causal effect.

What aspects of a given event can be said to be 'causally relevant' naturally depends on its nature and complexity, as well as the observer's interpretation and level of understanding of what is transpiring (e.g., compare watching any athletic competition with or without knowledge of its rules). That said, when looking at the stimuli used in Experiment 2, we can at least confidently classify the 'focal objects' (see § 3.1.2) as causally relevant. We were therefore able to reuse the videos from that previous experiment to test participants' memory of causally relevant details.

The main difference between Experiment 3 and Experiment 2 is that, in the task phase, participants now had to identify and reject *lure* items (in addition to making judgments about order). These lures were similar to actual still images from the videos they had seen, with only the change that the focal objects had switched positions. In cases where there was just one focal object present in an original still, the lure version would show the other focal object in its place. This manipulation ensured that the scenes produced in this fashion were not only unfamiliar to participants but also expressly incompatible with the causal structure of the events shown (which would not have been the case had background or peripheral elements been modified). The cued memory task employed otherwise resembled Experiment 1's, with 50 % of test items being authentic stills that could come before or after the cue image (corresponding to 'in-episode' items from Experiment 1), and the remaining half lures. We retained our temporal arrangement manipulation from Experiment 2.

It should be noted that the focal objects are causally relevant *within* a clip for all stimuli, even those that are fragmented, reversed or both. After all, every clip of every video features moving or colliding objects, or a combination of the two. What fragmented and reversed stimuli lack is a causal structure that bridges individual clips: Details about the state of the focal object in the first clip of a fragmented video stimulus are irrelevant to the events in subsequent clips, whereas in a coherent video, the states of the focal objects in the first and last clip are connected by an unbroken chain of causes and effects. Because the detailed state of the focal objects is in this sense more informative in coherent stimuli, we hypothesized that participants should be better at identifying lure items in the causally coherent condition.

4.1. Method

4.1.1. Participants

We recruited 88 adult participants for this experiment.[4] The sample size was chosen because power analyses of pilot data (Cohen's $d = 0.35$ in a $N = 9$ sample) indicated this number of participants would be required to reach >90 % power in a paired samples t-test comparing the mean score on forward coherent items with that on reversed coherent items – one of our preregistered analyses (https://osf.io/3hvs7/?view_only=2fa323f6c158432f82f52bf07a0b2741). Prospective subjects again had to be fluent in English and unfamiliar with the general study design. As pilot data suggested an increased number of low-quality, chance-level responses to this version of the task, we set more stringent exclusion criteria: Instead of removing only those from analysis who gave the same answer to every question item, we now also excluded those who showed no variation in two or more of the twelve question blocks. We further excluded participants who responded to five or more questions within 400 milliseconds or less. These criteria led us to discard the response data of nine participants and recruit replacements until we had reached the target $N = 88$. Participants received £4.5. The process of giving consent on Qualtrics and performing the task on Pavlovia was the same as in the first two experiments.

4.1.2. Materials and procedure

The presentation phase of Experiment 3 was almost identical to that of Experiment 2, with the biggest difference being that now all the videos had an equal number of clips (five). As described in the method section of Experiment 2, these were obtained by removing the first and last clip from the complete seven-clip video. We used truncated stimuli principally because they were balanced with regard to the number of focal objects present in each clip: The terminal clips 1 and 5 featured only one focal object each, while in the middle clip 2, 3 and 4 they were both present at the same time. Eleven of the twelve stimuli were videos we had previously used in Experiment 2; the remaining one turned out to be ill-suited for the creation of lures in the systematic manner we intended, so we created a novel video instead. Blocking and

[4] Due to experimenter error, we do not have demographic data for this experiment. However, it was recruited from the same population and with the same methods as Experiments 1 and 2, and within a matter of weeks of each other, so we expect that the demographic features of this sample should be highly similar to those of the other two experiments.

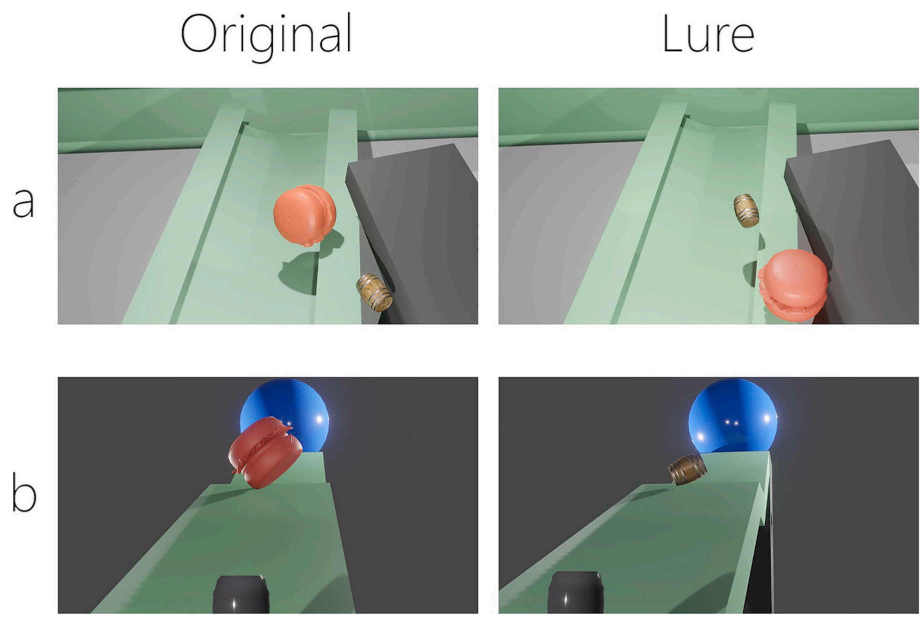

Fig. 6. **a** Lures created from original stills in which both focal objects were visible at the same time were created by simply switching their positions. **b** Example of original still with one focal object present compared to the corresponding lure. In the lure image, the focal object that would not have been visible (barrel) in this clip has been placed in the position of the other (red burger). Stimuli for Experiment 3 can be viewed at: https://osf.io/x5ckm/?view_only=509a59e553ec4acbbd4363b39212b371.

randomization during the presentation phase followed the design described in Experiment 2; assigning participants to one of two stimulus groups that flipped which half of the videos is shown in *forward* and which in *reversed* temporal arrangement was equivalent to the procedure employed in Experiment 1.

In the task phase, instead of just having to make a *before* or *after* judgement in a cued memory task, participants had three options and could additionally *reject* an item, i.e., report that it had never been presented. The mechanics of the cued memory tasks were otherwise the same as in Experiment 2 in that a still image from the central (in this case, third) clip of a video stimulus always served as a cue. There were twelve videos and hence twelve cuing images in total. The presentation of each cue marked the beginning of a question block comprising eight items, with the test items numbering 96 in total. There was an equal number of items from coherent and fragmented videos, as well as from forward and reversed videos (48 each).

In addition to causal status and temporal arrangement, we also manipulated the *authenticity* of image items. Half of the total items were *original*, which means they were unaltered stills from the video stimuli participants had seen. Consequently, rejecting them always constituted an error. The other half were *lure* images, which were modified stills similar to what participants had actually watched in the presentation phase. The correct answer to any lure item was always rejection. Lures were created by applying one of two rules: If both focal objects were present in the original image item, their positions in the lure would simply be switched (Fig. 6a). If an original image item featured just one focal item, it would be replaced with the other one that had not been present (Fig. 6b). Because every question block consisted of eight image items, presented consecutively and in randomized order, and half of the images were originals and the other half lures, participants would always be confronted with patently contradictory imagery within a block of test images (i.e., they would realize that they could not have seen both the original and the lure of a given image).

4.2. Results and discussion

To assess the interplay of our independent variables and their effect on memory performance we conducted a 2 (causal status) x 2 (temporal arrangement) x 2 (authenticity) repeated measures ANOVA. The dependent variable in this and all other subsequent analyses was the *accuracy* obtained in the cued memory task, i.e., the proportion of a participant's correct responses to a total of 96 test items (or subset thereof, when stated). All three factors – causal status, $F(1,87) = 72.2$, $p < 0.001$, $\eta_p^2 = 0.45$, temporal arrangement, $F(1,87) = 16.54$, $p < 0.001$, $\eta_p^2 = 0.16$, and authenticity, $F(1,87) = 37.65$, $p < 0.001$, $\eta_p^2 = 0.3$ – were found to have a significant effect on accuracy. In addition, we found significant interactions between causal status and temporal arrangement, $F(1,87) = 21.45$, $p < 0.001$, $\eta_p^2 = 0.2$, as well as causal status and authenticity, $F(1,87) = 4.23$, $p = 0.043$, $\eta_p^2 = 0.05$ (see Supplementary table 12 for all results).

We proceeded to analyze performance on items with *forward* temporal arrangement. We performed a 2 (causal status) x 2 (authenticity) repeated measures ANOVA, finding significant main effects for both causal status, $F(1,87) = 79.66$, $p < 0.001$, $\eta_p^2 = 0.48$, and authenticity, $F(1,87) = 40.02$, $p < 0.001$, $\eta_p^2 = 0.31$. The mean accuracy for coherent items was higher than for fragmented ones, and the mean accuracy for original items was higher than for lure items (Fig. 7).

To test whether participants were better at detecting and rejecting unfamiliar items when they related to an event with coherent causal structure, we performed a *t*-test comparing the mean accuracies for forward coherent lure items ($M = 0.51$, $SD = 0.27$) and forward fragmented lure items ($M = 0.35$, $SD = 0.22$). As we had hypothesized, participants were more likely to correctly reject lures for forward coherent videos than fragmented videos, $t(87) = 6.87$, $p < 0.001$, $d = 0.63$.

We further tested whether performance on forward coherent items ($M = 0.59$, $SD = 0.19$) was superior to that on reversed coherent items ($M = 0.48$, $SD = 0.17$) by means of another paired samples *t*-test. As in Experiment 2, we found this to be the case; $t(87) = 6.14$, $p < 0.001$, $d = 0.55$. Because it was a novel manipulation, we additionally tested this on just the subset of coherent lure items (scores ranging from 0 to 12). Forward coherent lure items ($M = 0.51$, $SD = 0.27$; as seen above) led to higher mean accuracy than reversed coherent lure items ($M = 0.44$, $SD = 0.26$), according to a paired samples *t*-test, $t(87) = 2.8$, $p = 0.006$,

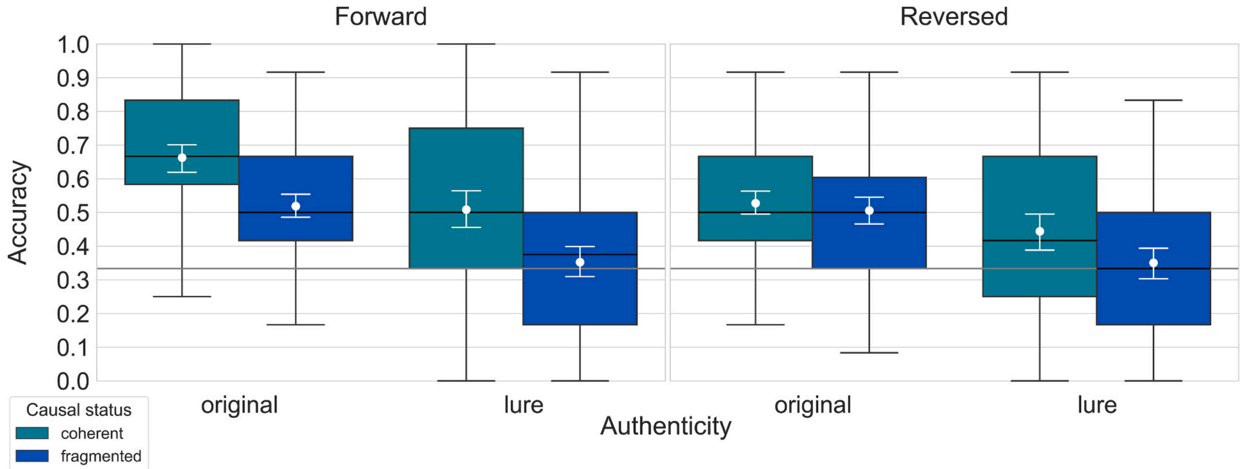

Fig. 7. Results of Experiment 3. Boxplots of accuracy across conditions. A repeated measures ANOVA showed that all three factors (causal status, temporal arrangement, authenticity) significantly affected accuracy; moreover, causal status significantly interacted with both temporal arrangement and authenticity. Considering the left panel (forward temporal arrangement) by itself, a 2 (causal status) x 2 (authenticity) repeated measures ANOVA revealed strong main effects for both factors on items with forward temporal arrangement. Boxes indicate quartiles, whiskers the spread of the data (1.5 x interquartile range), and the black line median accuracy. White dots represent mean accuracy; error bars indicate 95 % CIs. The horizontal gray line marks presumed chance-level performance (0.33).

though the effect size was markedly lower, $d = 0.24$.[5]

Taken together, Experiment 3 replicated the main finding of the preceding two experiments that causally coherent event structure leads to superior representation of order in episodic memory, and in addition shows that coherence also improves memory of causally relevant event details: Participants were significantly better at correctly rejecting lures for forward coherent items than any other item type.

5. General discussion

Across three experiments, we found causal coherence consistently affected recall of episodic memories in a cued memory task with dynamic video stimuli. Experiment 1 demonstrated that the order of three-clip videos of unfamiliar scenes was on average recalled better if the events presented were causally coherent. Experiment 2 extended this result to longer events, showing an improved recall of order in stimuli that were composed of five and seven clips. Experiment 2 furthermore used entirely novel video stimuli, indicating the results of Experiment 1 are not due to particularities of the material used. Finally, Experiment 3 replicated the effect of causal coherence on recall of event order seen in previous experiments and additionally showed that recall of causally relevant event details was better for coherent than fragmented events: When presented with ‘lure’ images that mimicked scenes they had actually seen but differed from them with regard to causally relevant details, participants in the coherent condition correctly rejected them more often.

Together, these experiments indicate that the episodic memory system is sensitive to the causal structure of dynamic events. The effect arose in an experimental setting in which the presentation of unfamiliar scenes was accompanied by no instructions beyond participants being asked to watch the material carefully. Importantly, control conditions in each of the three experiments demonstrated that causal coherence is fragile. Creating scrambled composites of different coherent stimuli (Experiment 1) or simply reversing their order (Experiment 2 and 3) led to recall performance that was indistinguishable from that of fragmented stimuli. As regards recall of causally relevant detail in Experiment 3, reversal of coherent stimuli led to a significant decline in the correct rejection of lures. Because the reversed controls were identical to the forward-oriented videos in every pixel, simply differently arranged, the effects we found cannot be explained by confounded visual characteristics of individual videos.

5.1. Order, segmentation and reordering

The most robust result obtained in these experiments is that causal coherence improves the accuracy with which event order is recalled. That explicit remembrance of causes and effects implies accurate recall of their order is almost definitionally assured. But what is not self-evident and our experiments demonstrate is that the mere observation of events induces the inference of causal relations to an extent that measurably impacts recall. *Order codes* (Friedman, 1993) have been proposed as a means of preserving important relational information about the sequence of past events without having to invoke absolute distance metrics. The three experiments suggest that coherent events are more likely to lead to the generation of an order code that specifies the relative temporal position of its parts to one another. Notably, this does not seem to be the result of a deliberate strategy on the part of participants upon noticing the causal structure manipulation: Across all three experiments, 69.4 % of participants chose to respond to optional debriefing questions that asked them to state what they believed the task had been about. Of the 150 participants who responded, only 9.2 % made implicit or explicit mention of causal relations.[6]

At this point we should ask: What is being ordered (or left disordered) here? After all, how the episodic memory system delimits individual episodes is an open question. Previous work on the subject indicates that temporal event boundaries in written narratives measurably impact strength of association between parts of a narrative in long-term memory (Ezzyat & Davachi, 2011). However, the clips whose order we asked participants to recall in our experiments probably do not correspond to the ‘natural’ units constituting an event’s representation in episodic memory. In our experimental tasks, participants had to make decisions about the order of temporal intervals that were evenly spaced in a continuous stream of visual input. Yet there is no reason to assume

[5] This final *t*-test was also the sole non-preregistered analysis detailed in this section.

[6] Some representative responses indicating unawareness of the manipulation: ‘It seems to me that the experiment was designed to investigate which shapes are easier for the human brain to remember.’ (Experiment 1); ‘No, I was simply confused and trying to keep up’ (Experiment 2); ‘To check if we could remember a sequence of events correctly.’ (Experiment 3)

that the units that make up an episodic memory are of an unvarying, quantized character; neither on a temporal, perceptual or even more abstract 'gist' level. What makes for an informative unit should be expected to depend on the overall structure of the particular event. Yates et al. (2023) recently made a similar argument relating to the definition (or lack thereof) of the term 'event' in cognitive science research. They suggested that instead of trying to pin down what is true of all event representations – trying to determine the general format and nature of an event – researchers should perhaps focus on how the content of a given event shapes its corresponding representation. It is arguably natural to assume that an event's *causal structure* plays an important role in this respect. Applying this reasoning to the domain of episodic memory, one may hypothesize that causal inflection points could serve as important segmentational cues that influence how a memory is subdivided into orderable units (with the potential consequence that representational granularity of an episode could mirror the granularity of inferred causal structure). Testing whether this is indeed the case could be accomplished by presenting dynamic stimuli *without cuts* to participants and probing whether their ability to correctly recall event order notably breaks down when they are asked to make 'before or after' judgments on timescales smaller than that on which the cause-effect structure manifests itself (e.g., analogous to using two images from different timepoints in the same clip in our segmented stimuli). This kind of experimental design may be something worth pursuing in future studies.

Another interesting order-related question concerns the reversal manipulation we employed in experiments 2 and 3: Why did playing the clips in reverse order mask their causal structure so effectively? Other recent studies investigating event memory and event perception employed naturalistic visual stimuli with a nonlinear temporal structure – such as the movies *Memento* (Antony et al., 2024), *500 Days of Summer* (Frisoni, Tosoni, Bufagna, & Sestieri, 2024), or a scrambled edit of a TV show episode (Grall, Equita, & Finn, 2023). In contrast to our own experiments, where reversal largely led to order recall equivalent to that of inherently disordered fragmented videos, the behavioral (Antony et al., 2024; Frisoni et al., 2024) and brain-imaging data (Grall et al., 2023) collected in these studies indicates that participants tended to (at least partially) infer the 'actual' temporal order, i.e., a timeline in which causes precede effects, from these scrambled modes of presentation. A somewhat similar phenomenon has been detected in the context of Michottian launching events, where participants under certain circumstances misreport the order in which colored blocks start moving (Bechlivanidis & Lagnado, 2016), making it compatible with a coherent sequence of causes and effects (A hits B hits C) rather than the unrelated and sudden motion onset actually presented to them (A hits B, C starts moving, B spontaneously stops).

Why did participants in our experiments show no similar tendency to reorder the reversed stimuli they were shown? First, our videos were much lengthier and visually more complex than those used by Bechlivanidis and Lagnado (2016), meaning that a legitimate mis*perception* of order (perceiving the reversed stimuli as temporally ordered) in this case would constitute a rather extreme illusion. Second, our videos showed unfamiliar events, i.e., situations for which no specific schemas exist, markedly distinguishing our design from the aforementioned experiments that used movies and TV shows as stimuli. It appears that even if an event is displayed nonlinearly, as soon as it contains sufficient cues to trigger appropriate schemas (narrative and otherwise), a reconstructive tendency sets in. Professional movies give participants a lot to work with: General knowledge (e.g. how people behave in a restaurant) and genre conventions (a romantic comedy probably will not feature extensive battle scenes) enable people to recognize deviations from the expected order of events; furthermore, complex narrative structure in general often goes hand-in-hand, and is arguably predicated on, rich causal structure – something studies using narratives as stimuli should perhaps take more consistently into account (as argued by Chen & Bornstein, 2024). Our stimuli featured identifiable causal relations but were devoid of narrative structure in any conventional sense and conformed to no overall predictable pattern; compared to movies they were 'inconsequential' events, in that they both began and ended arbitrarily. As such, their causal structure appears to be much more fragile. By contrast, classic work on the causal relatedness of parts of narratives illustrates that even fairly simple stories require rather intricate network representations to capture how sentences are connected causally (Trabasso & Sperry, 1985). Importantly, they also feature causal relations beyond mere mechanical interactions, for example 'psychological' and 'motivational' relations (Warren, Nicholas, & Trabasso, 1979), which potentially unfold on grander timescales than immediate physical force transfer. This also implies that to make sense of key moments in narratives, it is often necessary to relate the present to causally relevant (and potentially rather distant) past events – an expectation that appears to be borne out by findings reported in a recent neuroimaging study, which showed that grasping important narrative connections (an "aha" moment) is frequently preceded by the neural reinstatement of causally related past events (Song, Ke, Madhogarhia, Leong, & Rosenberg, 2025). It is possible that the strictly proximal and largely mechanical nature of the causal relations in our stimuli makes them less robust in the face of rearrangements than narratives, with their oftentimes more layered and intricate causal structure.

5.2. Content and construction

Our empirical findings were not restricted to effects on recall of order, but also touched on the fidelity with which particular content is remembered. In Experiment 3, we found that coherent events led to superior recall of causally relevant details. Although both the rough chronology (order) and the moving parts (causally relevant details) of an event must be registered to infer causal relations, memorizing the latter is to some extent a precondition of representing the former: A sequence of events could *in principle* be reconstructed by making the right educated guesses so long as enough of the actual content of these events has been retained; conversely, if no content is remembered, there is no basis for attempts at reconstruction.

When envisioning episodic recall as a more constructive process than a read-out of previously stored information (Schacter, Norman, & Koutstaal, 1998), it is clear how remembering important specifics of an event provides an advantage that goes beyond a small but quantifiable surplus of detail. Causal information is inherently relational, meaning additional content remembered of one part of an event will likely aid reconstruction of other, but connected, parts of the event. More relevant event details being available for coherent events would presumably result in constructive processes being more constrained when 'filling in' potential gaps than they would be for fragmented events. Indeed, it would seem that the reconceptualization episodic memory has undergone mainly in the last two decades, shifting the emphasis away from its function as an archive of past experiences and more towards the simulation of potential future (Schacter, Addis, & Buckner, 2007) or purely imagined or hypothetical experiences (De Brigard, 2014), is already highly suggestive of its interplay with causal cognition.

Nevertheless, influential accounts of how the episodic memory system (or wider neural networks hypothesized to overlap with or underpin its operations) may 'simulate' experience do not explicitly include causality (Buckner & Carroll, 2007; Hassabis & Maguire, 2007) or only mention it cursorily (Addis, 2020). Given its effect on recall, causal coherence should perhaps be considered as a potentially important factor in mental simulations of events. If certain kinds of imagining are conceived of as the rearrangement and repurposing of existing memory episodes to create a novel scenario (Schacter et al., 2007), it would be natural to expect that causally structured memory representations may provide more readily adjustable templates than those that are just sequences of perceptual material. It is probably harder to imagine prolonged arbitrary behavior or random impressions than events with a cause-effect structure. A system that aims to identify and represent the causal threads connecting parts of continuous experience may also gravitate towards using causal relations as a scaffold when generating potential future or entirely hypothetical scenarios.

Relatedly, it remains an open question whether our findings could be replicated in settings where people observe events with the (potentially *unwarranted*) expectation or belief that they are causally coherent. One might, to give a somewhat drastic example, present a cover story to participants that falsely implies that a stimulus they are about to see is governed by a causal pattern even though its dynamics could be described better in stochastic or associational terms. Would merely mentally imposing causal structure onto events, rather than inferring it from what is observed, make them more memorable? Although nothing seems to speak against it in principle, future empirical work will be required to determine whether it is enough for causality to be in the eye (or mind) of the beholder, rather than the phenomenon itself.

5.3. How does causal coherence provide structure to episodic memory?

Based on our finding that causal coherence leads to more accurate recall of event order, we have hypothesized that the representational structure of episodic memories may to some extent be a reflection of the causal structure identified while experiencing an event. We have also noted how the superior recall of causally relevant details could have implications for reconstructive memory processes. What has yet to be addressed is *how* a more durable representation of order may be achieved and *why* the episodic memory system seems to hold on to certain content of coherent events that is discarded in fragmented ones. There are at least three non-mutually-exclusive mechanisms which may account for this.

First, our results could be solely attributable to coherent events being more reconstructible. The results of our first experiment were arguably compatible with the notion that any memory advantages brought about by coherence are purely a recall phenomenon. Specifically, they were compatible with the proposal that an episodic memory is a ‘prediction in hindsight’ rather than a retrieval of the original event (Werning, 2020). The memory system may be thought of as generating these predictions by taking minimal event traces as input and outputting the most plausible guess of what occurred in a simulation-like format. Applying this more general view to our experiments, it could motivate the hypothesis that coherent events provide more cues for the post hoc reassembly of order and filling-in of forgotten or overlooked content and are therefore remembered more accurately. They need not be encoded any differently from fragmented events, and causal cognition would only be a factor in the recall phase. However, in Experiment 2 we found that accuracy of order recall for 5-clip videos and 7-clip videos was the same. If order recall is just a matter of informed guessing based on relevant cues, we might have expected a drop-off in performance in longer sequences that provided more opportunity for mistakes. This ties into a greater conceptual problem with suggesting variable ‘reconstructability’ as the sole explanation. Remembering envisioned as a piecing-together of traces does not address the issue of how traces are already grouped together before reconstruction begins. What seems to be taken for granted is that like is already paired with like to some degree: A ‘whole’ corresponding to a particular event of interest is presupposed, the challenge of recall made out to be just the ordering of its potentially jumbled components (traces of subevents). Without this assumption of pre-grouping, it is unclear how each instance of recalling a past experience avoids sorting all, or at least an untenable number of, autobiographical memories and memory fragments.

Second, an overt break of causal coherence can be viewed as providing strong evidence that tracking an event’s relations to subsequent events, or individual objects’ states in relation to other objects over time, will yield no useful information. There is evidence from previous studies that the informativeness of certain aspects of an event can be assessed (based on existing knowledge) while it is ongoing and that it predicts retrievability of both material that is more semantic (Schul & Burnstein, 1983) and more episodic in character (Huang, Velarde, Ma, & Baldassano, 2023). Because we hypothesize that causal relations are established postdictively, there will always be intervals during which something that recently happened is judged to be of uncertain causal informativeness. For instance, you may realize at work that you left the door to your home unlocked, but as long as no burglar comes by, it will be of no consequence; if, on the other hand, the house has been ransacked upon your return, the initial failure to lock it will be solidified as a causal node connected to many (presumably disagreeable) effects downstream. In other words, barring perfect predictability, only subsequent events reveal which events of the past graduate from pending to definitive causal relevance. This second potential mechanism can be described as causality gating what is encoded. It could in principle account for both worse order recall in fragmented items (‘It’s obvious that it doesn’t matter whether the jogger or the police officer came first.’) and less accurate recall of certain details (‘Remembering what the jogger was wearing won’t give me any clues about what the person pushing the stroller was wearing.’). A fragmented causal structure could therefore signal to the memory system that the order of the ongoing event is uninformative and there is no need to encode it precisely.

Third, coherence could shape encoding processes in a way that makes it more likely that order and certain content are retained. This would require some kind of on-line tracking of an event’s causal structure and entail the frequent retroactive modification of recently encoded memory. Once an effect is established (the house really was burglarized), the already existing memory of what is now confirmed to be a cause (neglecting to lock the door) is reactivated and updated. The memories of these events are no longer independent but a causal pair joined by a directed link. A neuroimaging study has provided evidence for the existence of a comparable process that led to the enduring updating of an existing autobiographical memory’s *affective* valence following an individual’s reflective reevaluation of its consequences (Speer, Ibrahim, Schiller, & Delgado, 2021). If causal relations were directly encoded into a memory episode’s representational fabric, order information would naturally emerge when recollecting either antecedent or consequent. This hypothesized difference in representation could also be described as coherent events being more *compressible* during encoding (and subsequent cycles of reconsolidation) than fragmented events, as causally linked events are treated as one ‘unit’ in memory rather than two (or more) separate ones.

As memory episodes almost never match the detail of immediate experience, their formation is an inherently compressive process in at least some respects. Beyond this basic observation, the phenomenon of hippocampal replay that can be consistently detected in sleeping rats provides a strikingly direct example of apparent memory compression: The rodents, during particular sleep phases, exhibit neuronal activity in hippocampal areas that seemingly ‘replays’ motion trajectories matching the movements they undertook while awake but proceeds several times more quickly (Lee & Wilson, 2002). These replay patterns were also detected in awake rodents, usually during states of comparative inactivity (Carr, Jadhav, & Frank, 2011). To what extent this accelerated revisiting of previously explored locations resembles episodic recall in humans is unclear. Nevertheless, subsequent behavioral studies with human participants showed that similar temporal compression occurs when remembering the path one took through a certain terrain (Bonasia, Blommesteyn, & Moscovitch, 2016) and that event segmentation predicts the degree to which navigational experiences will be compressed (Jeunehomme & D’Argembeau, 2020). Bonasia and colleagues note that they expect compression to be relevant beyond recall related to navigation and mention it may be affected by ‘the number of salient subevents.’ It is in this sense that we hypothesize coherent events to be more compressible. Although, prima facie, coherent events appear to be more intricate, with early parts affecting much later parts, they are also more redundant (from the cause one can link to the effect, from the effect one can link back to the cause). Compared to a set of stand-alone events (which together form a fragmented 'event'), a single coherent event of equal duration will always[7] be more compressible because its order information is distributed across the parts. Hence, order may be reconstructible even if details are missing from the memory representations of individual parts. On the flip side, the subset of details that are *necessary* to reconstruct that order (i.e., the objects involved and their changes over time) may be the backbone of an event's causal structure (see also Altmann & Ekves, 2019).

One attribute the second and third hypothesized mechanism share is that they posit the virtually uninterrupted tracking of causal connections between experienced events. In this view, we may say the mind's default assumption is that experienced events are usually causally coherent and that it accordingly continuously seeks relevant relations between parts of events. Because it helped to convey the reasoning underlying our experimental designs, in this paper we have mostly discussed causal coherence as something that is 'added' to a prototypical event and thereby enhances certain aspects of its representation in episodic memory. When considering real-life experiences, it may be more apt to think of it the other way around: Most events – especially those that we conceive of as consisting of various related parts – are causally coherent. Those that are not, are harder to remember.

The three accounts we have sketched here are not at all contradictory, in fact they are quite harmonious. Breaks in causal coherence may prevent the formation of postdictive connections linking effects to their causes, while causally coherent events that support those connections may enhance later reconstruction by teaching the 'simulation' system what effects follow from what causes. The present results cannot distinguish between these accounts or determine their relative importance, but future work can examine each of these potential mechanisms and their interactions to better understand how causal coherence shapes memory.

6. Conclusion

The continuous stream of experience abounds with too much perceptual information to warrant or allow perfect preservation in memory. The episodic memory system is therefore highly selective in choosing what details to hold on to and frequently abstains from recording the precise temporal sequence in which events occurred. The three experiments we described here showed that the causal structure of an event is often predictive of how well it will be remembered. Coherent events tend to produce more coherent memories which more accurately represent the temporal arrangement and certain perceptual details of the original experience. Whether these effects are due to easier reconstructability of coherent events, an ostentatious lack of causal structure signaling that certain information is 'not worth' being stored, events with causal structure being more compressible, or a combination of the above, remains an open question. Further research will also be needed to investigate whether the effect of causal structure persists in memories of longer events and perhaps is even more pronounced, as well as if it is modulated by an individual actively influencing an event rather than acting as a passive observer. For now, these findings highlight that the episodic memory system is prone to forgetting when made to serve as a vessel of disjointed sensory impressions but latches on to whatever causal patterns can be detected in the noise.

CRediT authorship contribution statement

Andreas Arslan: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonathan F. Kominsky:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

[7] This assertion rests on the assumption of all things besides causal structure being equal.

Declaration of competing interest

The authors have no conflict of interest to disclose.

Acknowledgements

The authors would like to thank Katharina Wenig for her assistance in recruiting participants and Gergely Csibra for discussions about theoretical considerations motivating these experiments. This research was partially funded by a CEU Research Support Scheme grant awarded to Jonathan F. Kominsky.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106317.

Data availability

The manuscript contains links to OSF repositories which contain the data we collected, the code we used in analyses, and scripts we used to run the experiments.

References

Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology, 11*(2), 233–259. https://doi.org/10.1007/s13164-020-00470-0

Altmann, G., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review, 126*(6), 817. https://doi.org/10.1037/rev0000154

Antony, J., Lozano, A., Dhoat, P., Chen, J., & Bennion, K. (2024). Causal and chronological relationships predict memory organization for nonlinear narratives. *Journal of Cognitive Neuroscience, 36*(11), 2368–2385. https://doi.org/10.1162/jocn_a_02216

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

Bechlivanidis, C., & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition, 146*, 58–66. https://doi.org/10.1016/j.cognition.2015.09.001

Bigelow, E. J., McCoy, J. P., & Ullman, T. D. (2023). Non-commitment in mental imagery. *Cognition, 238*, Article 105498. https://doi.org/10.1016/j.cognition.2023.105498

Black, J. B., & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior, 20*(3), 267–275. https://doi.org/10.1016/S0022-5371(81)90417-5

Bonasia, K., Blommesteyn, J., & Moscovitch, M. (2016). Memory and navigation: Compression of space varies with route length and turns. *Hippocampus, 26*(1), 9–12. https://doi.org/10.1002/hipo.22539

Bradski, G. (2000). The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer, 25*(11), 120–123.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, Article e9414. https://doi.org/10.7717/peerj.9414

Brownstein, A. L., & Read, S. J. (2007). Situation models and memory: The effects of temporal and causal information on recall sequence. *Memory, 15*(7), 730–745. https://doi.org/10.1080/09658210701539596

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nature Neuroscience, 14*(2), 147–153. https://doi.org/10.1038/nn.2732

Chen, J., & Bornstein, A. M. (2024). The causal structure and computational value of narratives. *Trends in Cognitive Sciences, 28*(8), 769–781. https://doi.org/10.1016/j.tics.2024.04.003

Clarke, J., & Tyler, B. (2024). Expecting the future: Entropy-based predictions in time perception. *Timing & Time Perception, 13*(2), 107–131. https://doi.org/10.1163/22134468-bja10105

De Brigard, F. (2014). Is memory for remembering? Recollection as a form of episodic hypothetical thinking. *Synthese, 191*, 155–185. https://doi.org/10.1007/s11229-013-0247-7

Exton-McGuinness, M. T., Lee, J. L., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research, 278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science, 22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin, 113*(1), 44. https://doi.org/10.1037/0033-2909.113.1.44

Frisoni, M., Tosoni, A., Bufagna, A., & Sestieri, C. (2024). Searching for linearity: Reconstructive processes reverse temporal scrambling in memory for movie scenes. Psychological Reports, 0(0). doi:https://doi.org/10.1177/00332941241282650.

Grall, C., Equita, J., & Finn, E. S. (2023). Neural unscrambling of temporal information during a nonlinear narrative. *Cerebral Cortex, 33*(11), 7001–7014. https://doi.org/10.1093/cercor/bhad015

Hanyu, N., Watanabe, K., & Kitazawa, S. (2023). Ready to detect a reversal of time's arrow: A psychophysical study using short video clips in daily scenes. *Royal Society Open Science, 10*(4), Article 230036. https://doi.org/10.1098/rsos.230036

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences, 11*(7), 299–306. https://doi.org/10.1016/j.tics.2007.05.001

Huang, J., Velarde, I., Ma, W. J., & Baldassano, C. (2023). Schema-based predictive eye movements support sequential memory encoding. *Elife, 12*, Article e82599. https://doi.org/10.7554/eLife.82599

Jeunehomme, O., & D'Argembeau, A. (2020). Event segmentation and the temporal compression of experience in episodic memory. *Psychological Research, 84*, 481–490. https://doi.org/10.1007/s00426-018-1047-y

Kominsky, J. F., Baker, L., Keil, F. C., & Strickland, B. (2021). Causality and continuity close the gaps in event representations. *Memory & Cognition, 49*, 518–531. https://doi.org/10.3758/s13421-020-01102-9

Lee, A. K., & Wilson, M. A. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron, 36*(6), 1183–1194. https://doi.org/10.1016/S0896-6273(02)01096-6

Lee, H., & Chen, J. (2022). Predicting memory from the network structure of naturalistic events. *Nature Communications, 13*(1), 4235. https://doi.org/10.1038/s41467-022-31965-2

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2020). Broken physics: A conjunction-fallacy effect in intuitive physical reasoning. *Psychological Science, 31*(12), 1602–1611. https://doi.org/10.1177/0956797620957610

Ludwin-Peery, E., Bramley, N. R., Davis, E., & Gureckis, T. M. (2021). Limits on simulation approaches in intuitive physics. *Cognitive Psychology, 127*, Article 101396. https://doi.org/10.1016/j.cogpsych.2021.101396

Michotte, A. (1963). *The perception of causality*. New York, NY: Basic Books.

O'Brien, E. J., & Myers, J. L. (1987). The role of causal connections in the retrieval of text. *Memory & Cognition, 15*(5), 419–427. https://doi.org/10.3758/BF03197731

Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. (1979). Children's comprehension and memory for stories. *Journal of Experimental Child Psychology, 28*(3), 379–403. https://doi.org/10.1016/0022-0965(79)90070-5

Radvansky, G. A., & Copeland, D. E. (2000). Functionality and spatial relations in memory and language. *Memory & Cognition, 28*, 987–992. https://doi.org/10.3758/BF03209346

Radvansky, G. A., Tamplin, A. K., Armendarez, J., & Thompson, A. N. (2014). Different kinds of causality in event cognition. *Discourse Processes, 51*(7), 601–618. https://doi.org/10.1080/0163853X.2014.903366

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8*(9), 657–661. https://doi.org/10.1038/nrn2213

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*(1), 289–318. https://doi.org/10.1146/annurev.psych.49.1.289

Schul, Y., & Burnstein, E. (1983). The informational basis of social judgments: Memory for informative and uninformative arguments. *Journal of Experimental Social Psychology, 19*(5), 422–433. https://doi.org/10.1016/0022-1031(83)90020-3

Shin, S. M., & Gerstenberg, T. (2023). Learning what matters: Causal abstraction in human inference. In *proceedings of the annual meeting of the cognitive science society* (Vol. 45, no. 45).

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning & Memory, 25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Sinclair, A. H., Manalili, G. M., Brunec, I. K., Adcock, R. A., & Barense, M. D. (2021). Prediction errors disrupt hippocampal representations and update episodic memories. *Proceedings of the National Academy of Sciences, 118*(51), Article e2117625118. https://doi.org/10.1073/pnas.2117625118

Singmann, H., Bolker, B., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., … Christensen, B. (2019). Afex: Analysis of factorial experiments (R package version 0.23-0) [computer software].

Song, H., Ke, J., Madhogarhia, R., Leong, Y. C., & Rosenberg, M. D. (2025). Cortical reinstatement of causally related events sparks narrative insights by updating neural representation patterns. *bioRxiv*. https://doi.org/10.1101/2025.03.12.642853, 2025–03.

Speer, M. E., Ibrahim, S., Schiller, D., & Delgado, M. R. (2021). Finding positive meaning in memories of negative events adaptively updates memory. *Nature Communications, 12*(1), 6601. https://doi.org/10.1038/s41467-021-26906-4

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605. https://doi.org/10.1037/0033-295X.99.4.605

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Strickland, B., & Keil, F. (2011). Event completion: Event based inferences distort memory in a matter of seconds. *Cognition, 121*(3), 409–415. https://doi.org/10.1016/j.cognition.2011.04.007

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language, 24*(5), 595–611. https://doi.org/10.1016/0749-596X(85)90048-8

Trabasso, T., & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24*(5), 612–630. https://doi.org/10.1016/0749-596X(85)90049-X

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York, NY: Academic Press.

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

Twardy, C. R., & Bingham, G. P. (2002). Causation, causal perception, and conservation laws. *Perception & Psychophysics, 64*(6), 956–968. https://doi.org/10.3758/BF03196799

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences, 21*(9), 649–665. https://doi.org/10.1016/j.tics.2017.05.012

Vallat, R. (2018). Pingouin: Statistics in Python. *J. Open source Softw.*, *3*(31), 1026 [computer software]. Doi:10.21105/joss.01026.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., & Van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wahlheim, C. N., Eisenberg, M. L., Stawarczyk, D., & Zacks, J. M. (2022). Understanding everyday events: Predictive-looking errors drive memory updating. *Psychological Science, 33*(5), 765–781. https://doi.org/10.1177/09567976211053596

Warren, W. H., Nicholas, D. W., & Trabasso, T. (1979). Event chains and inferences in understanding narratives. *New directions in discourse processing, 2*, 23–52.

Werning, M. (2020). Predicting the past from minimal traces: Episodic memory and its distinction from imagination and preservation. *Review of Philosophy and Psychology, 11*, 301–333. https://doi.org/10.1007/s13164-020-00471-z

Yates, T. S., Sherman, B. E., & Yousif, S. R. (2023). More than a moment: What does it mean to call something an 'event'? *Psychonomic Bulletin & Review, 1-16*. https://doi.org/10.3758/s13423-023-02311-4