Cognitive Systems Research 75 (2022) 25–35

Contents lists available at ScienceDirect

Cognitive Systems Research

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

Modeling human memory phenomena in a hybrid event memory system

David H. Ménager [a,*], Dongkyu Choi [b], Sarah K. Robins [c]

[a] Department of Electrical Engineering and Computer Science, University of Kansas, 1520 West 15th Street, Lawrence, KS 66045, USA
[b] Department of Social and Cognitive Computing, Institute of High Performance Computing, Agency for Science, Technology and Research, 1 Fusionopolis Way, Singapore 138632, Singapore
[c] Department of Philosophy, University of Kansas, 1445 Jayhawk Blvd., Lawrence, KS 66045, USA

ARTICLE INFO

Keywords:
Event memory
Episodic memory
Human memory phenomena
Human memory errors
Hybrid theory of event memory

ABSTRACT

Human event memory stores an individual's personal experiences and produces their recollections with varying degrees of accuracy. To model this capacity, we recently developed a hybrid event memory system that combines aspects of the two main theories proposed in the philosophical literature. We aim to model a complete range of human event memory phenomena – successful remembering, misremembering, and confabulation – using this framework. In this paper, we review our hybrid event memory system and present empirical results from a remembering experiment we conducted using this system. The results show that our system successfully models the full range of human event memory usage and errors.

1. Introduction

Event memory stores information acquired from experience, making it possible to remember those events once they have passed.[1] When humans exercise this capacity, retrieving past events from memory, the accuracy of their recollections varies widely. Attempts at remembering range from wholly accurate to wholly inaccurate, and often land somewhere in between. These variations in recollection accuracy are mostly imperceptible to the remembering subject. Psychologists who have documented this variation have engaged in an extensive investigation of the causes and consequences of false memories (e.g., Loftus, 1998; Schacter, 2019). In response, philosophers who theorize about memory have recently been engaged in a debate over how to taxonomize memory errors, identifying the requirements on remembering and the ways that distinct types of memory fall short of these requirements (Bernecker, 2017; Michaelian, 2016, 2020; Robins, 2016, 2019, 2020). Doing so generally involves a tripartite distinction between successful remembering, misremembering, and confabulation. The two most prominent accounts, the causal and the simulation theories, differ in terms of how they characterize successful remembering and how it differs from misremembering and confabulation. There is a general consensus about how to order these states: successful remembering is all or mostly accurate, misremembering involves a moderate amount of error, and confabulation involves the most extreme errors. Still, the views differ about how to best capture the nature and sources of these differences, and the debate over which philosophical position does best is at a stalemate. We believe that the lack of clear implementational details about either view contributes to this impasse because a complete evaluation of the theories is not yet possible. Our recent work (Ménager, Choi, & Robins, 2021) describes a novel hybrid theory of event memory that takes steps to address this issue. In this paper, we argue that our view accommodates and explains the full range of event memory phenomena by combining aspects of existing causal and simulation theories, and we further provide empirical evidence to support this claim. Importantly, we do this by moving beyond just a theoretical description. We provide a computational implementation, making it possible to directly assess whether or not our view provides an adequate account. We recognize that other computational models of event memory existed before (Laird, Lebiere, & Rosenbloom, 2017; Norman, Detre, & Polyn, 2008; Nuxoll & Laird, 2004; Rosenbloom, 2014; Sohn, Goode, Stenger, Jung, Carter, & Anderson, 2005), but up until now, these models have not included a discussion of how they categorize recollections with relation to false memory and the requirements for remembering.

The remainder of this paper is organized as follows. In Section 2, we introduce a taxonomy of event memory phenomena and discuss how the two existing philosophical theories account for them. Next, in Section 3, we discuss our hybrid theory and argue that it provides a broader coverage of event memory phenomena, when compared to existing views. In Section 4, we present empirical evidence showing our system's performance on a remembering task conducted in a simulated

* Corresponding author.
E-mail address: david.menager@parallaxresearch.org (D.H. Ménager).

[1] In this paper we opt for the term 'event memory' rather than 'episodic memory' so as not to provoke debate with those who consider episodic memory a uniquely human ability. See Rubin and Umanath (2015) for a discussion of this term.

https://doi.org/10.1016/j.cogsys.2022.05.003
Received 18 March 2021; Received in revised form 15 April 2022; Accepted 25 May 2022
Available online 2 June 2022

domain. During our analysis we show how the elicited recollections link back to our discussion of event memory phenomena. Then we discuss the experimental results further in Section 5 and present some future work in Section 6 before we conclude.

2. Existing theories of event memory phenomena

Philosophers and psychologists who study memory recognize a distinction between successful remembering and false memory. The former are accurate and the latter are not. Studies of false memory explore the factors that influence the production of these inaccuracies. In misinformation studies, participants are provided with misleading, inaccurate, or altered information about previous experiences. Researchers then observe the extent to which these lures are incorporated into participants' recall of those experiences. One of the most prominent misinformation techniques is the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), where participants are instructed to memorize a set of related items and are subsequently tested on whether other related items were also in the set. Suggestibility studies, in contrast, encourage participants to imagine various events, experiences, and activities in vivid detail. Over time, this activity leads some participants to claim to remember events derived from these imaginings, which never actually occurred (e.g., Loftus & Pickrell, 1995). These general techniques are extensively used and the results are well-replicated. In response, many urge a further distinction amongst false memories, based on the severity of the inaccuracy. False memories that often result from misinformation, and involve distortions and other mild inaccuracies, are misrememberings. Those that are wholly inaccurate, as happens in suggestibility studies, are confabulations. This research and general division guides philosophical theorizing about memory and its error states.

Amongst philosophical views of remembering, there is a general divide between causal and simulation theories. Causal theories (Bernecker, 2010, 2017; Debus, 2010; Robins, 2019, 2020) distinguish these three memory phenomena, remembering, misremembering, and confabulation, in terms of two features: (1) whether there is a memory trace, namely, an event-specific representation that derives from the prior event; and (2) whether its activity produces an accurate recollection. An experience of remembering that involves both of these features is an instance of successful remembering. Failures involving the first feature are confabulations and failures involving the second are misrememberings. Although causal theorists themselves do not elaborate much on the view's implementation, they give some hints as to what it would look like. One can, for instance, conceive of a system with stored event representations and a retrieval process that involves an interaction between these stored representations and the cues used to guide recall, perhaps along with other machinery. This sketch of an implementation illustrates how successful remembering would occur—namely, via the retrieval and reactivation of an event representation, or memory trace. Misremembering, in contrast, would involve a malfunction or disruption during that process, which would interfere with the retrieval or reactivation of the trace. The causal view does not, however, offer any guidance for how confabulations are produced. The only gestures at an implementation are in the direction of states that involve traces; nothing is said about what happens in cases where the representations are produced without traces, as is the case in confabulation.

Simulation theories, in contrast, have encouraged a broader taxonomy of memory states, including not only misremembering and confabulation, but veridical and falsidical forms of confabulation (Michaelian, 2016). Simulationists use two key features to divide up these states: accuracy and reliability. Unlike causal theorists, they do not appeal to memory traces. There may be such traces, at least on occasion, but they are not necessary and are thus irrelevant to the simulationist taxonomy. Instead, simulationists appeal to the reliability of the process by which the memory system generates representations. Accurate representations produced by reliable processes are instances of remembering. Failures of reliability produce confabulations, which are veridical if accurate and falsidical if not. When representations are produced by reliable processes, but fail to be accurate, it is a case of misremembering. Like the causal theorist, the simulationist does not have much to say about implementation. The system is likely to rely on more generalized and schematized knowledge, with little to no inclusion of traces and a focus on processes for combining the schematic content in reliable ways. Beyond these general suggestions, however, the simulationist does not provide any specific guidance about the internal parameters by which reliability will be produced or assessed.

The debate between causal and simulationist theories of remembering is ongoing. Theorists defending each position argue over whether a causal connection is necessary for remembering and which taxonomy best aligns with the range of forms of human event recollection can take. The implementational details of these theories have not featured much at all in these discussions. While the debate has led to increased clarification of each position and introduction of positions in the middle ground like (Werning, 2020), the issue of causation remains unresolved. Our aim in this paper is to move beyond this stalemate by evaluating these theories on other grounds. By shifting the focus to implementation, we have identified a novel and interesting frontier on which to assess the strengths and limitations of both of these approaches and our hybrid alternative. Since our hybrid theory has an implementation, we can ask how it carves up errors in the system and see whether it provides a clear account of event memory phenomena. We are particularly interested in questions of how confabulation comes about and how distinctions are made within the system between misremembering and confabulation. In the next section, we review our hybrid theory before continuing our discussion in this direction.

3. Hybrid theory of event memory

Our recent work (Ménager et al., 2021) introduced a hybrid theory of event memory that brings together important aspects of both the causal and the simulation theories, aiming to cover the entire range of human event memory phenomena. Additionally, our theory moves the debate over theories of memory in a new direction, because we implemented our hybrid theory in the context of a cognitive architecture (Choi & Langley, 2018), enabling the evaluation of its implementational commitments. In this section, we briefly review our hybrid theory of event memory and its associated implementation, and then introduce the extensions made for the purpose of the current work.

3.1. Theoretical assumptions

Existing theories of event memory commit to a single representational form to explain all event memory phenomena. In our previous work, we argued that problems accounting for the full range of event memory phenomena stem from this commitment, and further, that the problems of the causal and the simulation theories exhibited reciprocal strengths and weaknesses because they center on distinct phenomenon and corresponding representational form. The complementary nature of the two theories suggested that we could improve the theoretical account of memory by combining the representational commitments of both into a hybrid theory. In this hybrid view, we hypothesize that:

- Event memory is a long-term memory that stores episodes and schemas;
- Episodes are propositional representations of specific events;
- Schemas are first-order propositional templates with probabilistic annotations;
- Event memory elements are organized in hierarchies;
- Retrieval cues play a central role in remembering; and
- Remembering an event involves performing structural matching and probabilistic inference.

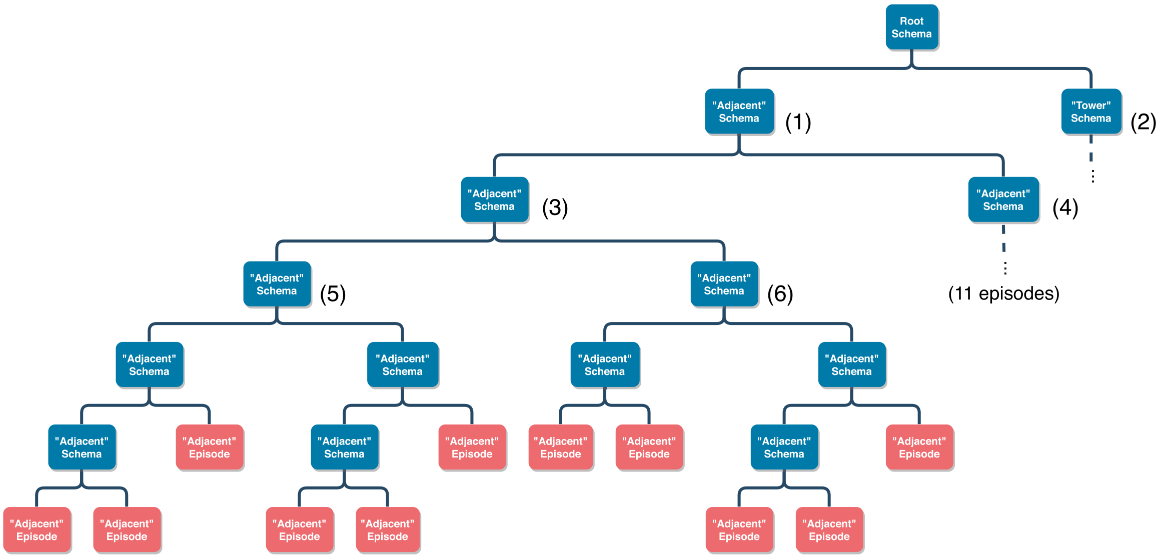


Fig. 1. A sample generalization tree for Blocks World from Ménager et al. (2021). The blue nodes except for (2) depict hierarchically organized schemas for `adjacent` class. The red leaf nodes are the specific episodes that belong to this class. The node (2) is the top-level schema for `tower` class, which has a similar structure underneath that is not shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

More specifically, our hybrid theory commits itself to the storage, maintenance, and use of episodes and schemas in a hierarchy. In making any explicit commitments at all, our theory goes beyond what has been provided by either causal or simulationist proposals. These philosophical accounts do not make any such explicit commitments, but the views have at least some architectural implications. That is, the theories are compatible with some implementations and not others. Ultimately, these theoretical accounts must be evaluated by their plausibility – i.e., whether an implementation of the theory is possible, and whether it produces the phenomena it should. Our approach begins from a commitment to basic implementational features implicated by each theory. Episodes correspond to the causal representation of events, while schemas generated probabilistically across a series of events correspond to the simulationist process. Episodic contents describe the remembering agent's external and internal states, which include perceived objects and their attributes, as well as hierarchical beliefs inferred from these percepts. Like in the causal theory, experienced events are operative in producing an episode and, in that sense, a causal link exists between a specific event and its episodic representation. In contrast, schemas are probabilistic summaries of episodes and other schemas. Rather than describing a specific event, schemas represent a range of possible outcomes and are therefore aligned with the simulationist approach to a process of constructing representations in a reliable manner. Both episodes and schemas are stored in a generalization hierarchy, as shown in Fig. 1, such that the leaf-level elements are episodes (shown in red color) while layers of schemas (shown in blue color) exist at higher levels. The schemas summarize their children by aggregating all their contents and storing probabilistic annotations with them. This event hierarchy forms a general-to-specific taxonomy in which the event memory elements are connected by IS-A links from a child node to its parent.

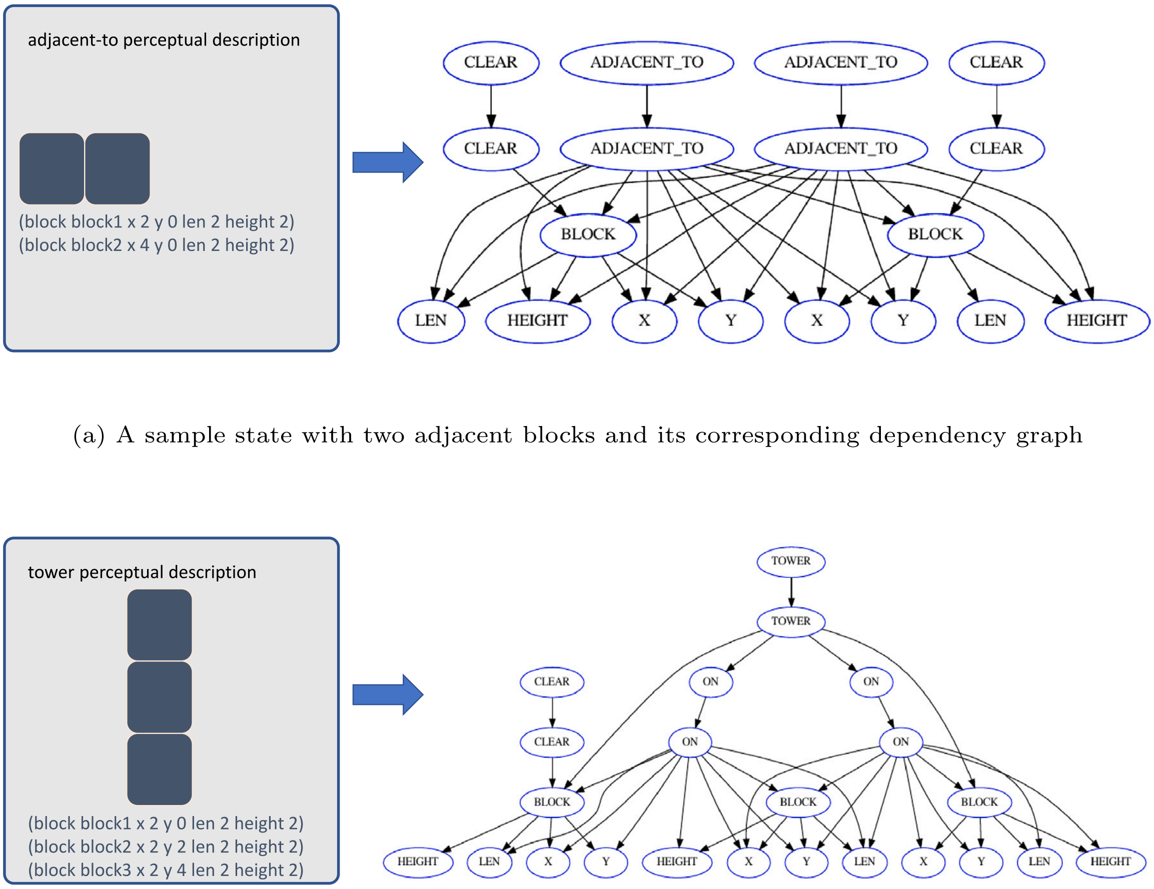
Furthermore, our theory assumes that remembering occurs in response to a retrieval cue. Given a cue, the event memory attempts to produce an event that is consistent with the cue contents. A structural matching process decides where in the event hierarchy the retrieval should happen, using a type of similarity metric. When retrieval happens from an episode stored in memory, remembering is a straightforward return of that episode. However, when the system deems that a schema is the best match, remembering an event involves probabilistic inference over the schema to collapse the probability distributions and produce a specific instance of that schema. This inference process is inherently approximate and can result in inaccurate recollection of events, providing ways to model human event memory errors.

3.2. Hybrid event memory system

We implemented our theory within the context of a cognitive architecture, enabling us to build event memory-enabled agents and test various aspects of our theory. The system encodes each episode or schema as a dependency graph like shown in Fig. 2, where we give two distinct examples from Blocks World. Fig. 2(a) depicts a situation where two blocks are adjacent to each other, touching at one side, while Fig. 2(b) shows three blocks stacked to form a vertical tower. In each case, the cognitive architecture represents perceived objects as typed predicates with attribute-value pairs. Our event memory system converts these perceptual inputs to trees of height 1, where the root node encodes the type of the perceived object (e.g., block) and the children nodes are its attributes (e.g., length, height, x position, and y position) connected to the root by a directed edge.[2] The system represents relations like `on` and `tower` defined using perceptual inputs or other relations as trees of height 1 over their component nodes. The root node of this tree holds the type of the relation, and edges connect it to the components that participate in the relation. We also added an additional generic node which covers all the possible definitions of a relation, because a relation may be defined disjunctively over different objects. In the examples shown, each of the relational definitions, `on`, `tower`, `clear`, and `adjacent`, exist under their respective generic root node with the same name. Finally, the nodes in an episode store the specific values of the state they encode, whereas nodes in schemas contain conditional probability distributions over possible values. In this manner, the system forms a Bayesian network for each schema.

As the agent encounters new situations in the world, our system inserts them as episodes into its event memory. During this process,

[2] A tree is a recursive graphical structure composed of nodes and edges whereby nodes are (un)labeled containers for data and edges define parent/child relationships among the nodes. A parent node has one or more outgoing edges to other nodes. Each node connected to a parent node is a child of that node. A child node may have one and only one parent. A root node is a node with no parent, and a leaf node is a node with no children. The root node is significant because it is the entry point of the data structure and all other nodes in the tree are ordered with respect to it. A tree is distinct from general graph structures because for any given node in the tree, there does not exist a path from this node back to itself. Finally, the height of a tree refers to the longest path from a leaf node to the root.


adjacent-to perceptual description
(block block1 x 2 y 0 len 2 height 2)
(block block2 x 4 y 0 len 2 height 2)
CLEAR
ADJACENT_TO
ADJACENT_TO
CLEAR
CLEAR
ADJACENT_TO
ADJACENT_TO
CLEAR
BLOCK
BLOCK
LEN
HEIGHT
X
Y
X
Y
LEN
HEIGHT
tower perceptual description
(block block1 x 2 y 0 len 2 height 2)
(block block2 x 2 y 2 len 2 height 2)
(block block3 x 2 y 4 len 2 height 2)
TOWER
TOWER
CLEAR
ON
ON
CLEAR
ON
ON
BLOCK
BLOCK
BLOCK
HEIGHT
LEN
X
Y
HEIGHT
X
Y
LEN
X
Y
LEN
HEIGHT

(a) A sample state with two adjacent blocks and its corresponding dependency graph

(b) A sample state with a tower of three blocks and its corresponding dependency graph

Fig. 2. Dependency graphs representing `adjacent` and `tower` states from Blocks World.

the system sorts the new episode through its event generalization hierarchy using a structural mapping procedure that attempts to match every node in the new episode to a corresponding node in an existing event memory element. The structural mapping procedure employs a similarity metric to compute the quality of match between the two elements. This metric is the Bayesian Information Criterion (Koller & Friedman, 2009) which measures how well the node in existing event memory element can explain the its counterpart in the new episode. Lower scores are better than higher ones. Hence, the system considers the lowest-cost match, summed across all the nodes, as the most similar element in the event hierarchy.

The insertion process generates an event hierarchy such that leaf-level elements are episodes, and layers of progressively generalized schema exist on top of them. Given an event hierarchy, like the one shown in Fig. 1, and given an episode, the insertion process starts from the root node of the hierarchy. From there, the event memory system utilizes its similarity metric and compares the match costs of the current node and its children with respect to the new episode being inserted. If the current node is the lowest-cost match among them, the new episode becomes a new child of this schema. The system then updates the schematic structure and the probability distributions accordingly and the insertion is complete. If, however, the lowest-cost match is one of the children, the system first updates the current schema to reflect the addition of the new episode and recursively moves to visit the lowest-cost child. During this recursive process, if the system reaches a leaf node and determines that the leaf node is the best match, the memory system schematizes this element to incorporate the new episode. This generated schema now stores probability distributions that cover both the previously existing and the new episodes.

Using the populated event memory, the system is capable of producing a recollection of an event in response to a retrieval cue. This involves a two-step process. First, the system finds an event memory element (an episode or a schema) in the generalization hierarchy that best matches the target event, through the same structural mapping procedure used during the insertion process. Then, it produces a fully instantiated event from the best matching element. If the retrieved element is an episode, producing an event instance is a trivial process because the episode itself is a fully instantiated event. If the retrieved element is a schema, however, the system must find a consistent set of variable assignments for the schema through probabilistic inference and constraint satisfaction problem solving before it can produce an event instance. The probabilistic inference outputs a set of constraints over the domain of possible values each variable can take, while the constraint satisfaction problem solving step produces the recollection by attempting to assign a specific value to each variable, minimizing the number of conflicting assignments in the system of variables.

In our previous work (Ménager et al., 2021), our event memory system required a complete solution from the constraint satisfaction step, thereby producing a fully consistent event at all times. If the system could not return a complete solution, it would simply return with failure, resulting in no recollection. Doing this, however, prevented the system from generating partially correct recollections, which is necessary to model human memory errors.

Since our current interest is in memory errors, we extended our system for the current work. Specifically, we did so by switching from our previous all-or-nothing constraint satisfaction procedure to a flexible one that can return partial solutions. This means that the system will always be able to generate a recollection, even if some of the variable assignments violate their respective constraints. In such cases, the variables with invalid assignments are not recollected. With this extension, the system is able to recollect partial states, so that when it cannot assign values to all the variables it will not simply fail to remember the event altogether. This, in turn, enables us to model incomplete or erroneous recollections humans often generate.

4. Experimental analysis

The goal of our current work is to demonstrate that our hybrid theory explains the full range of memory phenomena, including successful remembering, misremembering, and confabulation. To do this, we conducted an experiment in a simulated Blocks World. As we briefly described in Section 3.2, the agent perceives blocks and infers relations among them, generating a collection of perceived objects and the inferred beliefs at any given time as a state. The event memory system encodes such states as episodes and inserts them into memory. We used the experimental setup from our previous work (Ménager et al., 2021) as a starting point and extended it with several additional test measures to demonstrate our theory's broad coverage of event memory phenomena. We first presented a random sequence of state observations to the agent and then tested to see if it can remember those events when presented with a retrieval cue. Through this experiment, we aim to verify the following two hypotheses:

- Our event memory system shows three distinct groups of behavior, and
- These three groups map to the three human memory phenomena, successful remembering, misremembering, and confabulation.

In the rest of the section below, we first describe our experimental design in detail and explain the data generation process. We then discuss the clustering technique we used to verify the first hypothesis and present the result. After that, we show how those clusters map onto the philosophical account of event memory phenomena using decision tree analysis.

4.1. Experimental setup

For this experiment, we generated ten random sequences of 50 states in Blocks World. These states were drawn from two distinct classes of situations, shown in Fig. 2, with 50% probability for each. The first class, called `tower`, describes a scenario with three blocks arranged in a vertical tower. Situations that belong to the second class, called `adjacent`, contain two blocks placed adjacent to each other, touching on one side. When we drew samples from these classes, the configuration of the blocks was determined by the drawn class, but the dimensions, the positions, and the names of the blocks could vary. We provided each sequence of 50 states to our event memory system. For each sequence, we initialized the event memory to empty and let the system incrementally populate its memory as it encountered the events in the sequence.

Once the system finished storing all the events from the sequence, we generated retrieval cues by sequentially choosing each of the 50 states and taking 20 random subsets of the selected state according to a specified degree of completeness. To generate these 20 subsets, we randomly removed a portion of nodes and their incoming and outgoing edges from the corresponding full state until they met the specified completeness requirement. We provided these retrieval cues to the system and measured its recollection responses.

We considered this process as one epoch and repeated it ten times for each sequence. In every epoch, we modulated the completeness of the retrieval cue. We initially supplied full states as retrieval cues in the first epoch and gradually reduced their completeness by 10% as we moved to the subsequent epochs. By the tenth epoch, the completeness of retrieval cues drops to only 10% of the original states. This would result in a total of 100,000 recollection trials (10 sequences × 10 epochs × 50 states × 20 subsets as retrieval cues). After inspecting the generated partial-state retrieval cues, however, we realized that some trials used ambiguous retrieval cues, so we filtered them out before running the memory system over them. We labeled a cue ambiguous whenever its contents were a subset of both `tower` and `adjacent` classes. A good example of such a retrieval cue is `((block A) (block B) (on-table B))` because both classes can contain these elements. We removed these cues because it would not have been fair to ask the memory system to remember the correct event when the retrieval cue itself does not uniquely identify the target class. As a result of this filtering, we had 88,839 recollection trials available for the system.

4.2. Generating recollection trial data

As our system ran over these trials, we collected various measurements regarding the recollection performance of the system, capturing whether or not the retrieved memory element had the same structure as the target event specified in the cue, as well as other measures on the quality of the recollected event. We performed two analyses over this dataset as we describe in the next sections. Each data point from our recollection trials consists of several features: (1) recollection coverage; (2) recollection reverse coverage; (3) true positives; (4) false positives; and (5) false negatives. These are defined as:

Recollection coverage Percentage of nodes in the target event that are matched to nodes in the recollection. For example in Fig. 3, each node in the target event matches to a unique node in the recollection. The match only considers whether the nodes play the same role in their respective graphs, and does not check whether their assignments are equal. So, the recollection coverage is 100% in this case.

Recollection reverse coverage is the percentage of nodes in the retrieved event memory element that are mapped to nodes in the target event. Appealing again to the figure, in this case since the recollection contains two nodes unmatched in the target event, the recollection reverse coverage is 3/5. Both of these coverage measures only check whether the nodes are present in their respective event, rather than checking the value assigned to these nodes. Hence, these features in combination give a sense of whether the target event and the recollected event share the same structure.

True positives In our context, a true positive occurs when the value of a node in the recollection matches the value of its counterpart in the target event.

False positives False positives are cases where an element in the recollection does not match the target event. This can occur in two ways. In one case, a recollected event could contain a value-assigned node that is not present in the target event. In the other case, a false positive could occur when the recollected event contains a value-assigned node and the target event contains that same node, but with a different assigned value.

False negatives False negatives are cases where the recollected event lacks something that is included in the target event. False negatives can also occur in two ways. First, the recollected event could be missing a node that exists in the target event. Second, the recollected event could contain a node that is in the target event, but assign it a value that does not match its value in the target.

Our set of features does not include true negatives. This omission reflects the difficulty of quantifying what is not represented. In this research project, we make the closed world assumption: elements not present in either the target or recollected event are assumed false. This

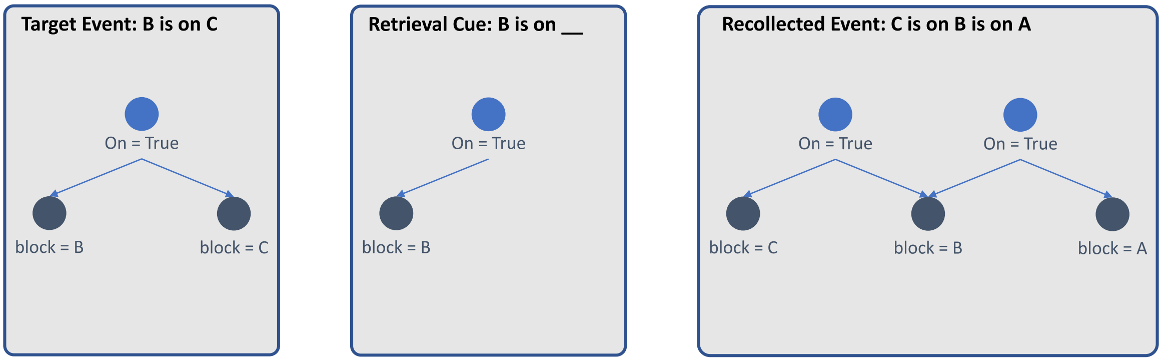


Fig. 3. Example target event with its associated retrieval cue and recollected event.

places an artificial limit on true negatives, when there are actually an unbounded number of items that could be false in both the target and recollected events. We can track true negatives in a limited sense, in cases where a node available in the recollection does not have an assigned value, indicating that the node could have been but is not present in the recollection, and also there is no corresponding node that matches it in the target event. Providing this abbreviated measure alongside true positives and false positives/negatives can result in misleading comparisons, so we have chosen not to include it.

4.3. Clustering recollection trial data

We first analyzed the recollection trial data to find any groups of data points that share unique characteristics. To do this, we needed to plot the data points in the multi-dimensional feature space and attempt to identify any meaningful clusters among them. Following this, we would need to let the unique characteristics of each cluster help us map it onto a memory phenomenon. But having five features to analyze in a dataset is challenging for many clustering algorithms. One common approach for dealing with multi-dimensional data is to reduce the high-dimensional feature set down to a lower dimensional space via feature selection or dimensionality reduction. A possible strategy for feature selection might be to replace a subset of the features with derivative measures that summarize the data, thereby reducing the total number of features to a more manageable number. In our case, we thought to summarize the performance measures by sensitivity and specificity. Doing so would result in four total features when accounting for the other two structural matching measures. Although a clustering algorithm could feasibly handle four-dimensional data, we unfortunately could not provide a reliable measure for specificity, which is defined as:

$$specificity = \frac{tn}{tn + fp}, \tag{1}$$

such that tn is the number of true negatives, and fp is the number of false positives. As explained above, because we make the closed world assumption, we cannot properly quantify true negatives and therefore cannot define specificity without introducing artifacts in the data. In this circumstance, our only remaining option was to employ Principal Component Analysis (PCA) to perform dimensionality reduction (Jolliffe, 2002). Doing so would facilitate our cluster analysis by removing noisy or non-predictive features in order to identify clusters in the data more easily. Using this method we reduced our five-dimensional feature set to a two-dimensional one. Fig. 4 shows the contribution of each original feature to the two principal components.

Examining the features reveals that the first component roughly gives equal weight to each component, having a slight negative bias against true positives and false positives. Component two, on the other hand, receives significant contributions from true positives and false positives, but puts less emphasis on the other features. These principal components account for about 98% of the variance in the data. Given these principal components, we plotted the data in Fig. 5.

We began the analysis with visual inspection. We saw one distinct cluster at the bottom right of the figure separated from a larger mass of data points on the left. Closer inspection of the latter revealed that the data points on the left split into two clusters near Component1 = 1 on the horizontal axis.

Although a visual inspection of this data gave us an intuitive sense of the clusters that exist in our data, we assigned each data point to a cluster in an analytical manner in order to understand how the event memory phenomena correspond to these clusters. We chose a clustering algorithm, OPTICS (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, Blondel, Prettenhofer, Weiss, Dubourg, Vanderplas, Passos, Cournapeau, Brucher, Perrot, & Duchesnay, 2011), to do this task. OPTICS is ideal for our dataset because it is a density-based clustering algorithm that can identify clusters of any shape, size, and density. Density-based clustering algorithms typically work best over two or three-dimensional data where data is more densely distributed. Other clustering strategies including Gaussian mixture models (Reynolds, 2009) or k-means (Lloyd, 1982) are less well-suited to approach our data because they make assumptions about the covariance structure of the data and can only discover Gaussian clusters. This results in clusters that resemble ellipses or circles, which is a limiting assumption.

OPTICS can also handle noisy data points that the other clustering algorithms cannot easily deal with. These characteristics are important in our context because the potential clusters in our data can vary in all of these aspects. Furthermore, tuning the OPTICS model only requires setting a couple of intuitive parameters[3] and, importantly, the number of desired clusters is not one of them.

Fig. 6 shows the discovered clusters in our recollection trial data. The scatter plot displays the data points assigned to each cluster with a different color. The result indicates that OPTICS identified three clusters in our data, shown in blue, green, and orange colors, respectively. There are also a number of noisy data points, shown in red color. The first cluster, Phenomenon 0, is in the left corner of the scatter plot and contains 78,443 data points.

Phenomenon 1 is the green cluster in between the other two and contains 52 members. Lastly, Phenomenon 2 is to the left side containing 10,336 data points. The remaining eight data points are labeled as noise. The plots near the component axes display the density of each cluster projected along the principal components. This clustering

[3] We carried out a grid search for the optimal parameter settings for `eps` and `min_samples`. The `eps` parameter sets the maximum Minkowski distance between two data points in the same neighborhood, and `min_samples` sets the minimum number of required data points in the neighborhood of a point in order for it to be a core node of a cluster. This parameter was set as a fraction of the total number of samples in the data. The grid search yielded an `eps` of .001 and the `min_samples` of 7×10^{-5}. We also set the cluster extraction method to 'dbscan'.

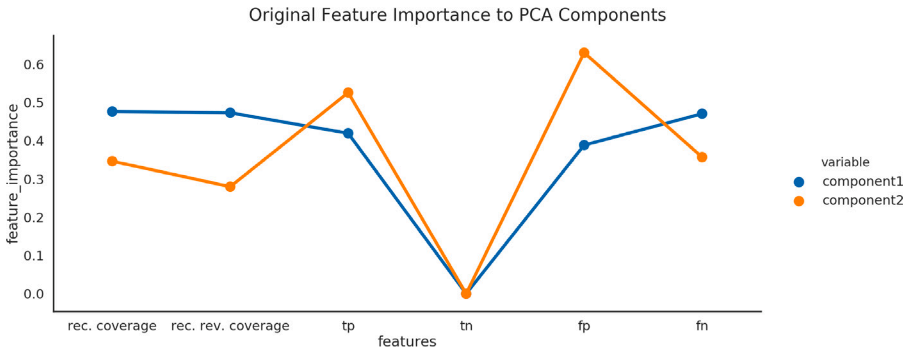


Fig. 4. Contributions of original features to principal components.

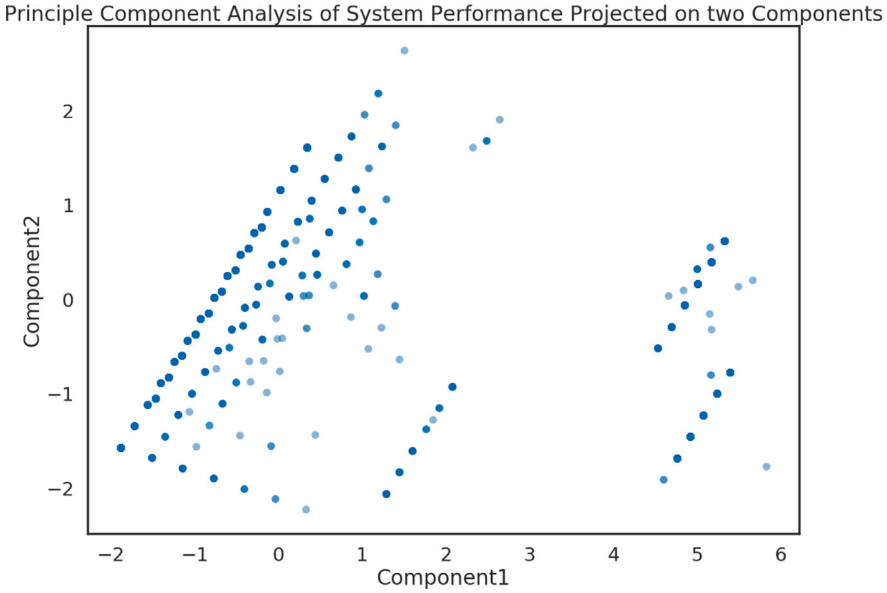


Fig. 5. Recollection performance data scattered across principal components.

result received a silhouette coefficient of .613 in a range from $[-1, 1]$. A score of 1 is the best and indicates that the points in a cluster are close together while being far from points in other clusters. A score of 0 indicates overlapping clusters, and negative values suggest that points have been assigned to wrong clusters. In summary, these clustering results show that each cluster is well-defined and separated from the others. Our next objective is to describe these clusters in terms of their performance characteristics.

4.4. *Mapping clusters onto event memory phenomena*

Discovering the three clusters in our experimental data is an encouraging sign, since we hypothesized that our system can model three kinds of human event memory phenomena. But, our analysis up to this point only shows that there are three groups of data points and does not characterize these clusters. To obtain such information, we augmented the data points in our original dataset with their cluster assignments and generated Table 1 summarizing each cluster in terms of the recollection performance.

From this elaboration on each cluster, we can now attempt to characterize the three corresponding recollection phenomena. The table shows that Phenomenon 0 has near perfect recollection coverage and perfect reverse coverage. Additionally, this phenomenon produces, on average, 14 true positives, and six false positives. This indicates that the members of this cluster are high-quality recollections that cover the target event without adding extraneous information. Also, recollections in this group are rarely wrong about specific details about items in the state. In such erroneous cases, the target event may have a block of height 10, but the system remembers the height as 15.

Next, Phenomenon 1 achieves 61.11% recollection coverage and has perfect reverse coverage. This implies that recollections in this cluster recover much of the structure from the target event, but still fail to capture it completely. Looking at the other performance measures, this cluster contains recollections that show elevated amounts of false positives and negatives when compared to Phenomenon 0, but still lower than Phenomenon 2. These moderate values suggest that Phenomena 1 captures middle-of-the-road recollections that contain a mixture of accurate and erroneous responses, both in terms of details and structural quality.

Finally, the recollection coverage and the recollection reverse coverage for Phenomenon 2 at 55.05% and 63.79%, respectively, suggest that the retrieved event memory element and the target event do not share the same structure. This implies that the event memory system remembered the wrong kind of experience in response to the cue. This is also reflected in its low true positives (4.17), and high false positives (14.82) and false negatives (10.01) scores.

Having described the phenomenal characteristics of each cluster, we now ask how these clusters map onto the phenomenal categories that researchers use to classify types of event memory phenomena. There might be a number of ways to construct a mapping from cluster to event memory phenomenon, but our analysis above suggests that one plausible assignment is to label Phenomenon 0 as successful remembering, Phenomenon 1 as misremembering, and Phenomenon 2 as

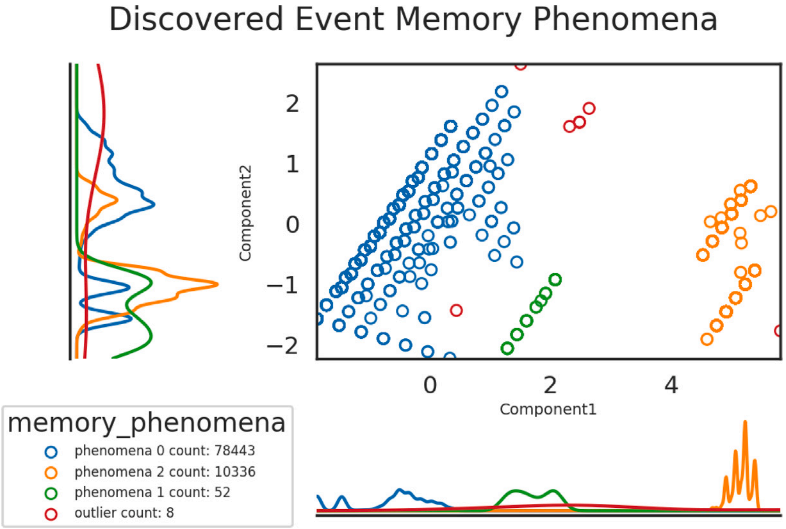


Fig. 6. Recollection performance data annotated with cluster assignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Average cluster performance characteristics.

True positives		False negatives		Rec. Rev. Cvg.	
Phen. 0	14.09 (4.20)	Phen. 0	0.01 (0.13)	Phen. 0	100.00% (0.18)
Phen. 1	12.40 (1.98)	Phen. 1	7.00 (0.00)	Phen. 1	100.00% (0.00)
Phen. 2	4.17 (1.11)	Phen. 2	10.01 (2.00)	Phen. 2	63.79% (5.78)
False positives		Rec. Cvg.			
Phen. 0	6.08 (4.43)	Phen. 0	99.97% (0.56)		
Phen. 1	5.58 (1.97)	Phen. 1	61.11% (0.00)		
Phen. 2	14.82 (2.60)	Phen. 2	55.05% (5.78)		

confabulation. With this mapping in mind, we can interpret Fig. 6 as showing the clusters for successful remembering and misremembering close to each other on the left and having the confabulation cluster further away at the bottom right corner of the scatter plot. Also, most of the recollections are instances of successful remembering, followed by confabulation, and then misremembering with the least frequent occurrences in the data.

While we believe this is the most straightforward way to do the mapping in both studies, it is not perfect. For example, when we checked whether the clusters, as we have labeled them, truly reflect the phenomena we are trying to model, a concern arose pertaining to the false positives. Recall that we assigned Phenomenon 0 to successful remembering despite containing 6.08 false positives. Is it acceptable to include six false positives in successful remembering? Without a baseline implementation from previous work in the form of an implemented human event memory system, it is difficult for us to judge this. One might have hoped that successful remembering would involve no false positives, but this seems to be an unreasonable standard. Studies of human memory often determine success based on whether key features of an experience are retained. Without knowing precisely what is encoded and how much information is available to be encoded, it is difficult to determine what proportion is retained in cases of success.

For this first implementation, we are willing to accept the six false positives in successful remembering because our results show that this phenomenal category is sufficiently distinct from the other clusters. We suspect that the number of false positives may be attributed to the limited amount of variations we could produce in the events our system experienced in the Blocks World. The system may have had trouble identifying the target event amongst very similar events, resulting in this result. Despite this limitation, we are still encouraged by the mapping we found between our clusters and the widely accepted categories of event memory phenomena. We believe that our results provide an important baseline for further research in this area.

5. Discussion of mapped memory phenomena in our system

In the previous section, we presented results that suggest our event memory system models the full range of human event memory phenomena. We did this by demonstrating the extent to which performance measures, like true positives and false positives, relate to known categories of memory phenomena. Given this encouraging outcome, we are further interested in providing an account of how each phenomenon arises in our system. We do this by showing how internal system parameters correspond to our discovered phenomena. In this stage of our analysis, there are four pertinent parameters:

Weighted distance: quantifies the quality of match between the cue and the retrieved event memory element. This parameter was z-score normalized and ranges from [−1.18, 3.47].

Retrieved element count: indicates the number of episodes summarized by the retrieved event memory element thus giving a measure of stability for the retrieved event memory representation. Stable event memory elements summarize many episodes and less stable elements summarize fewer. This parameter was z-score normalized from [−0.26, 23.32].

Retrieved element depth: describes the depth of the retrieved event memory element from the root of the tree. This parameter was z-score normalized and ranges from [−3.4, 3.97].

Retrieved element type: specifies whether the retrieved event memory element was an episode or a schema.

To conduct our analysis, we first annotated each record in our dataset to indicate which event memory phenomenon it was. Then, we dropped all performance measure features from the dataset, leaving

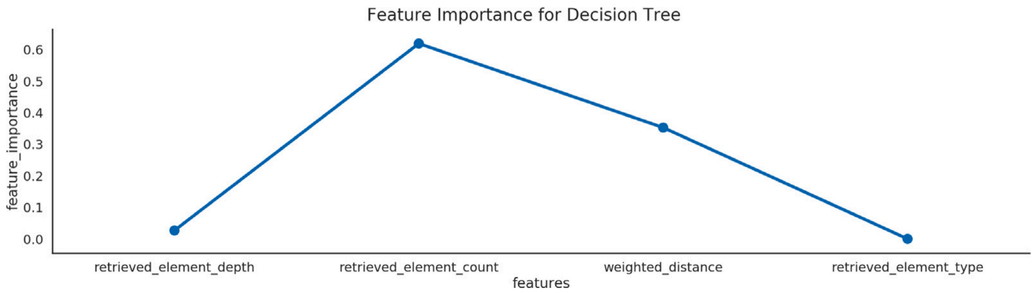


Fig. 7. Feature importance for decision tree trained over PCA'd data.

us a dataset containing only the four system parameters identified above. We applied a supervised machine learning technique, called a decision tree (Pedregosa et al., 2011), to discover the conditions under which our event memory system produced the different event memory phenomena. We balanced the class weights of the decision tree which ensured that it would not use the relative frequency of the phenomena to bias the predictions toward a particular class. To learn this decision tree, we split our dataset into a training and testing set, where the training data comprised 80% of the data. We then conducted 10-fold cross-validation over the training set and obtained 93.28% average accuracy and 93.37% accuracy on test, indicating that the model generalized well. We show in Fig. 7 that the decision tree relies heavily on the weighted distance and retrieved element count to distinguish the three classes.

Given these features, we plotted in Fig. 8 the training data points annotated with their class assignments along those identified features. We see the cluster assignments allowed the decision tree to obtain good separation in the event memory model's parameter space. The resultant decision tree trained over this data was quite large so we were not able to visualize the complete tree here, but Fig. 9 shows the top portion of the learned decision tree. Although we do not display the entire tree the not shown class assignments in lower-level nodes do not vary much.

Each node in the tree contains fields for the condition, entropy score, number of samples, value, and class. The condition field shows a Boolean condition on a specific event memory system parameter, and the entropy score is a measure of node purity. An entropy score of zero means that the node contains examples of one and only one phenomenon. Entropy scores greater than zero indicate that the node contains some mixture of phenomena. The mixture proportions in the node are subsequently shown in the value array. The samples field shows the total number of records in the dataset under consideration at that node. The class field indicates the class in which most of the examples satisfying the condition belong, and the color and intensity of the node match the predicted class and purity of that node. For example, the root node is clear because all the classes are equally probable. The adjacent node along the 'false' branch is colored green because most of the records there are cases of misremembering. Finally, we truncated the rest of the tree in order to present the most essential information in the figure.

The rules in the decision tree are conjunctions of conditions and can be discovered by sequentially evaluating the condition in each node, beginning from the root. If the condition evaluates to 'true', then the next condition to evaluate is found at the adjacent node along the 'true' branch. Otherwise, the next condition of the rule is the node along the corresponding 'false' branch. The complete decision rule constitutes any path starting from the root of the tree down to the leaf.

With this in mind, we examined the decision rules in the tree to obtain our theory's account of event memory phenomena. The first rule states that successful remembering occurs whenever the retrieved event memory element count and the cost of matching to the cue are low. A low event memory element count means that the retrieved representation was either an episode or a schema that summarized a small number of episodes. Low distance means that the quality of match between the cue and the retrieved element was very good. In other words, a route to successful remembering in our model was one where there were few possible options for retrieval and those options were well-matched to the cue.

The next decision rule states that confabulation occurs in cases where the retrieved element count is low, but the weighted distance is not low.

As for the previous rule, a low count means that the retrieved representation was either an episode or lower-level schema. The difference arises because of the difference in weighted distance. Confabulation occurs when there is a lack of high-quality matching between this small set of representations available for retrieval. This could occur because the retrieval cue is poorly specified. If it is incomplete or missing information, then retrieval may only identify a few possible representations, and each candidiate branch may seem to be a roughly equally bad match to the cue, causing the system to traverse to a path toward a highly unrelated representation.

The third rule shows another path to successful remembering, one that occurs whenever the weighted distance is low, but the retrieved event count is not. In these cases, the cue activates representations summarizing many episodes—i.e., a schema. Given the low weighted distance, there is enough matching between the cue and the schema to produce successful remembering.

The last decision rule, which predicts misremembering, occurs when both the retrieved element count and weighted distance are not low. That is, misremembering occurs in cases where the retrieved representation was a schema, but the schema was not particularly well matched to the cue. Higher-level schemas with larger retrieved element counts will often summarize over a range of events, which may differ enough to introduce errors in identifying the best match for the cue.

This decision tree demonstrates our theory's ability to account for different memory phenomena, and highlights the hybrid nature of our model. Two decision rules, the first and third, produce successful remembering. The first indicates successful remembering from episodes. The third shows successful remembering from stable schemas. Since our view is a hybrid theory, we expected to observe two routes of success in our analysis. One is consistent with causal accounts, which require the retrieval and activation of a stored representation. Our analysis shows that this is a route by which successful remembering is often produced, when the cue is well-matched to a small number of stored representations. The other is consistent with the simulationist account. The procedures our system uses to retrieve memory elements is presumably reliable by simulationist standards. Our analysis shows that such a reliable process yields successful remembering when the cues are connected to well-articulated and well-matched schemas. What is particularly significant about our analysis is the demonstration of how a single system can produce successful remembering of both forms from within a single architecture.

Our analysis also distinguishes both of these forms of successful remembering from confabulation and misremembering, and importantly, confabulation and misremembering from one another. Both forms of memory error occurred in cases where the weighted distance was comparatively high. This makes sense: errors are more likely to occur

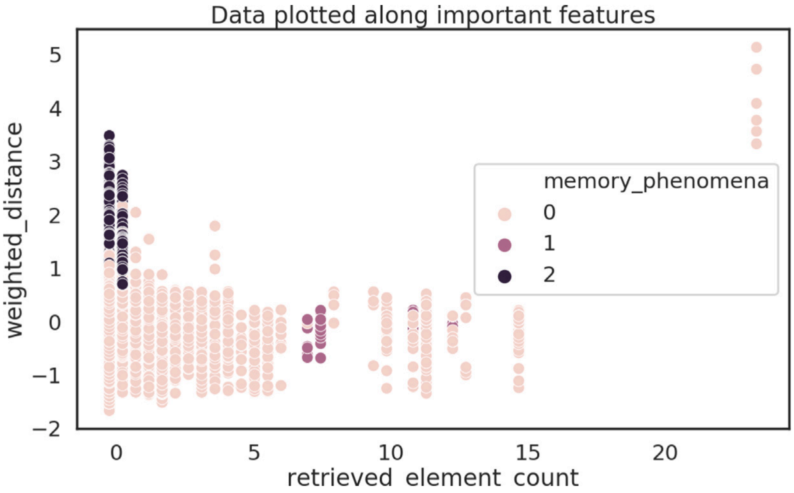


Fig. 8. Plotted data with labels from clusters found in PCA'd data.

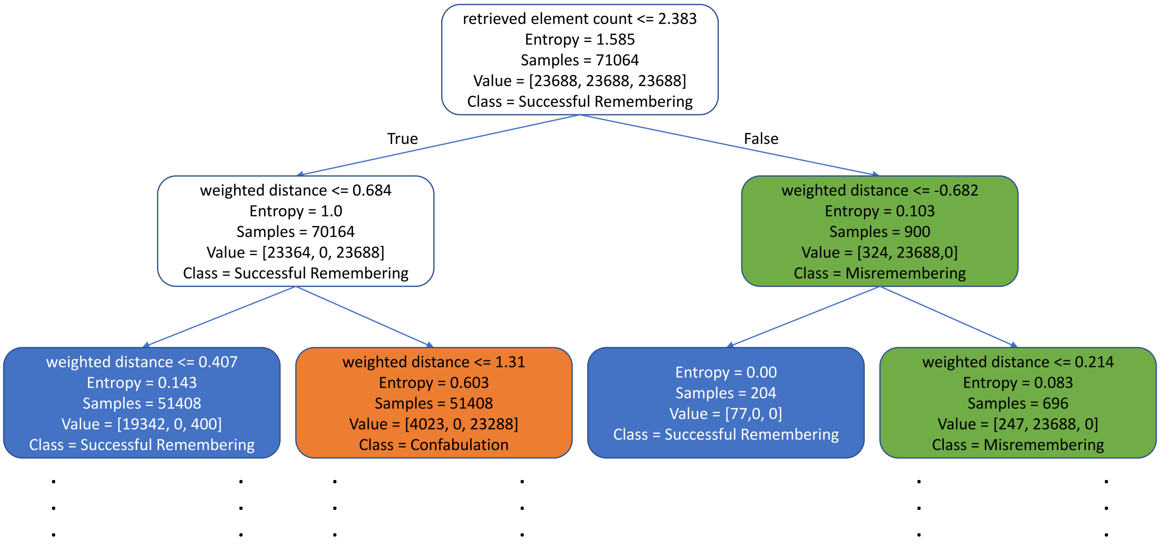


Fig. 9. First three levels of the learned decision tree. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

when the cue and possible items for retrieval are not well-matched. The difference between the two forms of error seemed to concern the kinds of representation involved in the mismatch. In our model, confabulation was more likely to occur in cases that were episode-driven (low retrieved element count), whereas misremembering was most likely to occur in cases that were schema-driven (high retrieved element count). Very little is known about the details of the cognitive and neural mechanisms by which these forms of errors are produced in human event memory, so a comparison on these grounds is not possible. Our results do, however, appear complementary to the experimental techniques by which these forms of error are induced, as discussed in Section 2. Misremembering is generally produced by misinformation studies, which induce errors via the presentation of related, but irrelevant information. This suggests that such errors are produced by confusions that result from the incorporation of schema-consistent information. Confabulation is produced by suggestibility studies, which induce errors via the creation of unique, imagined episodes.

These results and interpretations are preliminary, but they suggest a novel and integrative approach to explaining the forms of successful remembering promoted by both causal and simulation theorists, while also characterizing and distinguishing amongst forms of false memory.

6. Future work

As discussed so far, the experimental results showed that our system demonstrated three kinds of recollection outcomes which, through our analysis, we linked to the successful remembering, misremembering and confabulation. We additionally showed our theory's explanation of the conditions under which each phenomenon was produced. These results lead us to claim that our hybrid theory of event memory provides a plausible account of human event memory and our implementation successfully models its phenomena.

Given these promising results, we plan to continue our work along three main dimensions. First, we would like to capture the temporal dimension in our event memory representations. Currently, our system only handles state-of-affair representations, which describe perceived objects and relations defined over them. Many events, however, unfold over time and our system's inability to capture this is a serious limitation that we need to overcome. Doing so will open the door for us to build even richer event memory-enabled intelligent agents that can talk about their past, make predictions about their future, or even understand the goals and intentions of other agents.

Consequently, our second goal is to apply this extended system to plan, activity, and intention recognition problems (Ménager, Choi, Floyd, Task, & Aha, 2017; Mirsky, Stern, Gal, & Kalech, 2018; Pei, Yunde Jia, & Zhu, 2011). Traditionally, work in this area has focused on recognizing the top-level goals and intentions, as well as low-level actions of an agent. This, while interesting, may not be very useful for interactive collaborative agents. These systems will need to infer not only top-level goal and intention information, but also predict the specific methods used to complete tasks. Toward our goal of building such agents, we would like to extend our system to store hierarchical, temporal event representations which would allow it to not only infer the top-level goal of an observed agent, but also all the intermediate subgoals and intentions associated with them.

Both of these extensions rely on the underlying probabilistic inference system. Currently, this system treats real numbers categorically, instead of performing inference over real-valued variables. Hence, our third goal is to extend our system to handle both continuous and discrete probability distributions during inference. We believe this will require a hybrid Bayesian Network (Koller & Friedman, 2009) that should allow our system to operate in real-world environments where reasoning over numerical information is important.

7. Conclusions

Human recollections of events exhibit a range of event memory phenomena. Previous theories attempting to explain this have done so with partial success. In this work, we presented a novel hybrid theory of event memory which, we argued, explains the full spectrum of event memory phenomena. To support this claim, we conducted experiments in Blocks World and showed that our system accounts for both the successes and failures of human event memory usage. Although continued research in this area is necessary, we believe that our work provides a more complete model of the memory phenomena and represents an important step toward a complete understanding of human event memory.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is partially supported by A*STAR under its Human–Robot Collaborative AI for Advanced Manufacturing and Engineering (Award A18A2b0046). The first author was supported by Self Graduate Fellowship at the University of Kansas while performing this research. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the A*STAR or the University of Kansas.

References

Bernecker, S. (2010). *Memory: a philosophical study*. Oxford University Press.

Bernecker, S. (2017). A causal theory of mnemonic confabulation. *Frontiers in Psychology*, *8*, 1207.

Choi, D., & Langley, P. (2018). Evolution of the ICARUS cognitive architecture. *Cognitive Systems Research*, *48*, 25–38.

Debus, D. (2010). Accounting for epistemic relevance: A new problem for the causal theory of memory. *American Philosophical Quarterly*, *47*(1), 17–29.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*(3), 305–312.

Jolliffe, I. (2002). *Principal component analysis*. Verlag, NY: Springer.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT Press.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13–26.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.

Loftus, E. F. (1998). Illusions of memory. *Proceedings of the American Philosophical Society*, *142*(1), 60–73.

Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals*, *25*(12), 720–725.

Ménager, D. H., Choi, D., Floyd, M. W., Task, C., & Aha, D. W. (2017). Dynamic goal recognition using windowed action sequences. In: Workshops at the thirty-first AAAI conference on artificial intelligence.

Ménager, D. H., Choi, D., & Robins, S. K. (2021). A hybrid theory of event memory. *Minds and Machines*, 1–30.

Michaelian, K. (2016). Confabulating, misremembering, relearning: The simulation theory of memory and unsuccessful remembering. *Frontiers in Psychology*, *7*, 1857.

Michaelian, K. (2020). Confabulating as unreliable imagining: In defence of the simulationist account of unsuccessful remembering. *Topoi*, *39*(1), 133–148.

Mirsky, R., Stern, R., Gal, K., & Kalech, M. (2018). Sequential plan recognition: An iterative approach to disambiguating between hypotheses. *Artificial Intelligence*, *260*, 51–73.

Norman, K. A., Detre, G., & Polyn, S. M. (2008). *Computational models of episodic memory*. Cambridge University Press.

Nuxoll, A., & Laird, J. E. (2004). A cognitive model of episodic memory integrated with a general cognitive architecture. In *Proceedings of the sixth international conference on cognitive modeling* (pp. 220–225). Citeseer.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pei, M., Yunde Jia, & Zhu, S. (2011). Parsing video events with goal inference and intent prediction. In *2011 International conference on computer vision* (pp. 487–494). http://dx.doi.org/10.1109/ICCV.2011.6126279.

Reynolds, D. (2009). Gaussian mixture models. In S. Z. Li, & A. Jain (Eds.), *Encyclopedia of biometrics* (pp. 659–663). Boston, MA: Springer US, http://dx.doi.org/10.1007/978-0-387-73003-5_196.

Robins, S. K. (2016). Misremembering. *Philosophical Psychology*, *29*(3), 432–447.

Robins, S. K. (2019). Confabulation and constructive memory. *Synthese*, *196*(6), 2135–2151.

Robins, S. (2020). Mnemonic confabulation. *Topoi*, *39*(1), 121–132.

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

Rosenbloom, P. (2014). Deconstructing episodic memory and learning in Sigma. In: Proceedings of the annual meeting of the cognitive science society, vol. 36.

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review*, *122*(1), 1–23.

Schacter, D. L. (2019). Implicit memory, constructive memory, and imagining the future: A career perspective. *Perspectives on Psychological Science*, *14*(2), 256–272.

Sohn, M.-H., Goode, A., Stenger, V. A., Jung, K.-J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage*, *25*(1), 21–33.

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.