Published in final edited form as:
Trends Cogn Sci. 2017 December ; 21(12): 962–980. doi:10.1016/j.tics.2017.08.005.

Constructing experience: Event models from perception to action

Lauren L. Richmond and **Jeffrey M. Zacks**
Department of Psychological and Brain Sciences, Washington University in St. Louis

Abstract

Mental representations of everyday experience are rich, structured, and multimodal. In this article, we consider the adaptive pressures that led to humans' construction of such representations, arguing that structured event representations enable cognitive systems to more effectively predict the trajectory of naturalistic everyday activity. We propose an account of how cortical systems and the hippocampus interact to construct, maintain, and update event representations. This analysis throws light on recent research on story comprehension, event segmentation, episodic memory, and action planning. It also suggests how the growing science base can be deployed to diagnose impairments in event perception and memory, and to improve memory for everyday events.

Keywords

event cognition; episodic memory; event segmentation; action planning; binding

Representations enabling the efficient prediction of behavior

The last two decades have seen the emergence of a vibrant and robust body of research on *event cognition*—how people perceive, remember, think about, and respond to events [1]. One consistent conclusion from studies of event perception, memory, and action planning is that cognition uses structured representations of events, called *event models*, to capture information about the spatio-temporal framework, entities and objects, and other salient features of a situation. In particular, humans appear to maintain event models representing the current "state of play" of their immediate environment, called *working models*. Working models are actively-maintained, multimodal working memory representations that are stable most of the time but are updated at boundaries between perceptually-identified events. In a 2008 TiCS paper, Kurby and Zacks summarized studies that to that point had provided evidence for the existence and nature of working models, and began to characterize the mechanisms of their construction and updating [2]. Here, we address a fundamental question about these representations: Why does the brain have them? As cognitive representations go, event models are elaborate and resource-intensive. What unique capacities do they enable?

---

[*] Correspondence: jzacks@wustl.edu, 1 Brookings Drive, Campus Box 1125, St. Louis, MO 63130, USA.

We will give a speculative answer to these questions, integrating some recent developments in the field of event cognition.

Our hypothesis starts from the observation that adaptive behavior is predictive. Organisms across the phylogenetic spectrum anticipate how their environments will change and how their behaviors will affect those changes. Even the simplest predictions are of dramatic adaptive benefit—a fish that darts under a rock in response to a bird-shaped shadow is more likely to survive predation, and a bee that flies to plants whose color signals pollen is more likely to feed. However, humans (and other species) demonstrate predictive processing that is qualitatively more complex and subtle than these sorts of associations. Suppose you come across a friend in the middle of changing a bike tire. After a brief look at the scene, you can probably predict all kinds of things about the situation: that there is a bike pump or CO2 inflator around, even if none is visible; that your friend will put the wheel back on the bike; that your friend would respond positively if you offered assistance. We make detailed, subtle predictions about everyday events constantly and often do so with little awareness; predictions are a powerful feature of human cognition. Moreover, the brain appears to spend a substantial amount of its energy budget on predictive processing that puts it in a position to respond to predicted environmental features [3]. Thus, a re-representation that makes such predictions more accurate, more efficient, or both may be a worthwhile investment for the brain.

Psychologists have been interested in predictive mechanisms since Pavlov characterized classical conditioning [4], and in recent years there has been intense interest in predictive mechanisms in neuroscience [3, 5, 6]. One thing that is striking about these literatures is how powerful simple learning mechanisms can be—prediction has been observed in diverse domains including classical conditioning [4], visual feature detection [7] motion perception [8], motor control [9], scene perception [10], and language comprehension [11]. A common feature of models of prediction, across domains, is that they learn from experience, tuning their predictions based on a history of sequences in which cues or actions are followed by consequences [12]. We can view what these models learn as a representation of the statistical structure of a particular domain—motions of arms or distributions of objects in a scene, for example. It is often helpful to think about such structures as being shapes in a mathematical space. For example, observed arm movements could be represented in terms of the positions of joints, in terms of the angles between joints, or in terms of muscle torques [13]. A given action is a trajectory in a space whose dimensions are positions, angles, or torques. By observing a large number of reaching actions, a model can learn to anticipate based on the recent history of the arm's location in the space where it is likely to be a moment later. For reaching, the dimensionality of the space is modest; for other problems, as in the predation example given earlier, it may be much larger. However, whether the dimensionality is large or small, the structures that models of these domains learn tend to be *smooth*. The trajectory that is predicted does not jump wildly from one location to another. Many aspects of behavior and other natural phenomena are characterized by smooth dynamics. If one wants to predict how a reaching motion will unfold, smooth dynamics are a good bet. Research in motor control and embodied cognition has shown that learning systems can learn shapes in the relevant perceptual and motor spaces that allow for prediction by extrapolation. This sort of prediction-by-smooth-extrapolation is effective for performing perceptually-guided motor behavior [14] and can also be used offline for imagery and reasoning [15].

However, many aspects of everyday activity are *not* smooth. Suppose one observes Rebecca and Zach sitting across a table, and Rebecca is passing a carton of milk to Zach (see Figure 1 for an illustration of this scenario). If one were to represent their movements in terms of muscle torques, it would be very difficult to predict when Zach's arm would transition from resting to reaching; this onset is discontinuous, and challenging to relate to the other variables. More broadly, intentional agents in complex everyday environments produce many sequences that are very jumpy in the space of the relevant perceptual and motor variables. These sorts of functions are difficult for systems to learn and represent.

It is to address this challenge, we think, that event models evolved. *An event model re-represents a situation in a space that is nonlinearly transformed from the spaces of the sensory and motor variables, yielding structures for typical everyday activities that afford efficient learning and robust extrapolation*. An event model is not an arbitrary fabrication of the mind— to be useful, it needs to be closely coupled to the evolving state of affairs in the real world [16]. However, it is a re-representation of the world that alters its computational affordances. In our milk example, if one represents the situation not in terms of muscle torques but in terms of agents, objects, and intentional actions, then the transition from Rebecca's offering to Zach's accepting is smooth. The cost, here defined as 'computational effort' of building, maintaining and updating event models is the cost of going from a representation in the space of modality-specific perceptual and motor variables to a multimodal space of distal objects and entities. We propose that the answer to the question "Why does the brain have event models?" is this: in the transformed space of an event model, the trajectories of everyday activities are smooth and learnable.

In the following sections we will first review some historical foundations for research on event models. Then, we turn to current work in event cognition and its connections to other areas of cognitive science.

From situation models to event models

A substantial influence on current event model research has been older work in discourse processing on situation models [17–19]. These studies showed that both relatively deliberate and conscious inference mechanisms and fast and automatic predictive mechanisms contribute to text comprehension. For example, readers will sometimes draw predictive inferences from phrases such as "suddenly the actress fell," allowing faster processing of a phrase such as "The actress was pronounced dead" [20]. However, similar effects can also be produced by faster, simpler mechanisms such as semantic spreading activation and associative retrieval [21].

Situation models help the reader to make predictions about relations between elements in a text [17]. Bransford and colleagues asked participants to listen to sentences such as: 1) "Three turtles rested *beside* a floating log, and a fish swam beneath them" or 2) "Three turtles rested *on* a floating log, and a fish swam beneath them". Next, they were asked whether they recognized having read "Three turtles rested (*on/beside)* a floating log, and a fish swam beneath *it*." Those individuals who heard sentence 2 were more likely to falsely recognize "swam beneath it," and Bransford and colleagues suggested that this was because they were more likely to have predicted that the fish was beneath the log based on their model of the situation described by the text. People often make similar types of predictions about spatial relations among objects. For example, given the input "The book is to the right of the clock. The glass is on top of the book" one can easily predict that the glass is also to the right of the clock [22] even though this spatial relation between the glass and the clock was not directly stated.

Individuals also can use knowledge of semantic relations between items to drive predictions about plausible action plans and the objects likely to be involved in that action plan. Altmann and Kamide [23] designed a clever paradigm to investigate this, in which individuals were shown still images and listened to sentences that pertained to the images while their gaze location was tracked. For example, a participant might be shown a scene of a child sitting on the floor surrounded by a motorized train, a toy car, a ball and a cake. Participants might hear "The boy will move the cake" or "The boy will eat the cake". They found that individuals fixated on relevant objects prior to the object being mentioned when the spoken verb (e.g., 'eat') has a strong association with only one of the objects (cake) present in the scene [23, 24].

Situation models were originally described as representations constructed by readers of a narrative text, but similar considerations have been found to apply in the context of the comprehension of movies or interactive events. The term *event model*, thus, is a more general one referring both to event representations constructed during text comprehension and to representations formed from movies or live experiences [1]. Much of the recent research on event models has focused on the perception and remembering of visual events, though important studies of events in text comprehension continue. Quantitative research on event model construction from real life is still exceedingly rare—in part because live events cannot be repeated for multiple test sessions. However, two lines of evidence suggest that mediated events are a good proxy for live ones. First, as we have just seen, many manipulations have a similar effect on event cognition in reading and in movie viewing. This convergence across two kinds of mediated events with very different surface properties leads us to suspect that the effects might also generalize to unmediated events. Second, laboratory measures of event comprehension predict actual action performance [25].

Recent research on scene construction, episodic memory, and episodic future thought has highlighted the central role of event representations in memory and reasoning. Recalling something from the past or imagining something in the future appears to trigger a common scene construction mechanism, resulting in a structured mental representation complete with spatio-temporal context and vivid semantic and sensory details [26, 27]. Many of these same elements are present in building, maintaining and updating working models to re-represent the dynamics of the physical world. This suggests that a common representational medium for constructing event representations may be critical for ongoing comprehension, memory retrieval, and thinking about future or potential events. One suggestive finding consistent with this hypothesis concerns *boundary extension*, the phenomenon that people often remember pictures as having contained more of the world than they actually did [28]. People with amnesia do not show boundary extension, even when tested immediately following study [29]. This suggests that a common dysfunction can affect both online scene construction and episodic memory.

The link between remembering past events and imagining potential futures has led to the suggestion that one primary function of memory might be to predict and plan for the future [30], possibly by projecting one's self into a possible future [31]. If so, to what extent do these models re-use perceptual and motor systems for online perception and action control to do offline computations for memory and planning? This is a central concern of the current interest in *embodied cognition*; Box 1 discusses the role of embodied content in event models.

Segmentation

The segmentation of ongoing activity into meaningful events is importantly a matter of phenomenology: people spontaneously experience ongoing activity as consisting of meaningful events punctuated by boundaries between events [for an overview, see 2]. This subjective experience corresponds with phasic activity throughout the cortex [42], and can be easily and reproducibly captured by asking people to mark off the boundaries as they occur [43, 44]. Viewers of movies and readers of stories identify boundaries when more features of a situation are changing [45–48]. Choosing placement for event boundaries that tends to match those of other observers has downstream consequences for memory: the better an individual's segmentation matches that of the group, the better memory for the event is [49, 50].

Studies of the brain's phasic response at event boundaries have provided evidence that some resource-intensive activity takes place at the points where readers and viewers identify event boundaries [42, 51, 52]. Recently, interest has turned to identifying and characterizing the neural dynamics of event representations themselves. One approach has focused on the cortex. Hasson, Yang, Vallines, Heeger, and Rubin [53], using a creative movie-scrambling manipulation, characterized the duration over which representations in various parts of the cortex were sensitive to prior context. They discovered a hierarchy of *temporal receptive windows*, such that brain regions close to the sensory surfaces appeared to maintain information only for short durations, whereas association areas in parietal and frontal cortex maintained information for much longer durations. Baldassano and colleagues [54] used a data-driven method to identify points in time where the activity in a brain area shifted from one stable pattern to another. They found evidence for such regions throughout the brain, and found that their transition points corresponded with viewers' subjective event boundaries. Different regions transitioned on different timescales, in a way that corresponded well with the temporal receptive windows identified by Hasson and colleagues [53].

Another approach to the neural dynamics of event representations has focused on the hippocampus and surrounding structures. Building on work on place cells in the hippocampus [55], and on work on hippocampal representations of temporal context in human verbal memory [56, 57], Eichenbaum and colleagues have used electrophysiology in rodents to characterize *time cells*, which respond to specific timepoints within a temporal sequence [56]. Hsieh, Gruber, Jenkins, and Ranganath [58] trained humans on a picture memory task in which participants viewed a stream of pictures that contained repeating sequences, and individual pictures appeared at predictable locations within a sequence. Using fMRI, they found that hippocampal representations coded for specific objects in specific learned temporal positions, but not for object identity or temporal position alone. Further, they found that the pattern of activity in the left hippocampus changed more at boundaries between sequences than at picture-to-picture transitions within sequences (see also [54].

To investigate the causal link between the neural activity observed in fMRI studies and segmentation ability, segmentation was tested in a group of patients with brain injuries from the Vietnam head injury study [59]. Individuals with penetrating traumatic brain injury (pTBI) had lower segmentation performance, and those with larger lesions were especially impaired. A strong association between lesion size and coarse-grained segmentation was observed. However, there was little regional specificity with respect to specific aspects of impairment. For example, it was predicted that lesions in the ventromedial prefrontal cortex would result in poorer scores on videos that contained social information compared to nonsocial videos; this hypothesis was not supported [60].

In naturalistic activity, people are not simply passive observers but are also actors. Do people segment their own activities in the way as they segment the actions of others? Researchers have tended to implicitly assume the answer is “yes,” but little empirical research has tested this hypothesis. There is good reason to think that actors’ and observers’ representations of an activity often differ substantially, because actors and observers have access to different features of the activity [61, 62]. As an actor, one often has better access to the state of one’s plans and goals; as an observer, one often has better access to the larger physical context of action. Does this affect segmentation? A recent study from Swallow and colleagues suggests not: No differences were observed in segmentation of the event based on vantage point [63].

Viewpoint becomes a particularly acute issue when one considers the immensely important affective dimension of experience. Barrett and colleagues have proposed a theory of emotional experience that sees affect as a component of a predictive internal model that has much in common with working models as described here [64, 65]. A key feature of this account is the role of interoception—registering of one’s internal states—in constructing this internal model. When perceiving emotions, actors have access to interoceptive features, cognitions, and goals not available to observers, but observers have access to contextual features and facial expressions that may be less available to actors. Verbal labels can play an important role in aligning the features available to actors and observers [66]. An important question for future research is how the updating of affective aspects of internal models relates to the updating of event representations in actors and observers. Are they temporally coupled? Are they determined by the same mechanisms?

In short, people parse ongoing streams of activity into events, characterized by stable neural states that transition discretely at event boundaries. (See Box 2 for a discussion of structures richer than simple segmentation.) People segment others' activity as they observe it, and recent research suggests that they segment their own activity using similar mechanisms.

Prediction error-based updating

For working models to be helpful, they need to be updated at appropriate times. Failure to update leads to perseveration on a situation that has now changed; too-frequent and too-infrequent updating render working models unable to retain relevant situational features; and updating in the wrong places leads to working models that are a mishmash of information that is relevant and irrelevant, accurate and inaccurate. An event model system that perseverates, that jumps around constantly, or that fails to carve activity at its joints cannot facilitate smooth prediction. Thus, the mechanism by which working models are updated is crucial. One account of this mechanism is given by *event segmentation theory* [EST; 70] . EST proposes that working models are updated in response to transient increases in prediction error (see Box 3). As we have noted, a well-fitting working model improves a cognitive system's ability to make predictions about the near future. However, when things change in the world, the current working model's contents will become out of date, leading to errors in prediction. At such moments, it is adaptive to update one's working model. EST proposes that this updating is regulated by monitoring ongoing prediction error. When prediction error spikes, the current working model is flushed and information is gated into a new working model from (a) currently activated sensory and perceptual representations, (b) activation carried over from the previous working model, (c) associative retrieval from episodic memory, and (d) associative retrieval from semantic memory (see [2], Box 1).

Binding features into event models

When a working model is updated, some mechanism needs to bind together disparate features into a coherent representation in a space that will afford smooth prediction. In the example in Figure 1, simply representing “R holding milk” as an isolated feature does not afford smooth prediction; it is the multidimensional representation of the current milk holder, the state of the coffee cups, and so forth that makes prediction feasible.

The problem was articulated clearly by Baddeley in his description of the *episodic buffer* [92], a construct that has much in common with the working model construct we describe here. As Baddeley noted, such a representation probably leverages long-term memory to boost its capacity and durability [see also 93]. Baddeley suggested that one natural place to look in the brain for such representations is the PFC [92]; see also [70]. This possibility fits with neurological studies identifying action disorganization impairments with lesions to the PFC [94] and also with classical monkey and human physiology studies identifying sustained firing in the lateral PFC with working memory maintenance [95–98].

However, a growing body of evidence suggests that the hippocampus (HPC) and adjacent structures also play a critical role in rapidly binding disparate features into coherent complexes. The HPC has classically been conceived as a structure specialized for long-term episodic memory, spatial memory, or some combination of the two. However, a large and growing body of evidence supports the idea that the HPC is important for binding together disparate features of a situation [99–101], even when the retention interval is essentially zero [99–103]. For example, when people with HPC lesions were asked to view simple arrays of objects and then reconstruct them after a 4 s delay, they were disproportionately likely to swap the positions of two objects, indicating that they successfully retrieved the objects' locations, but had failed to bind individual objects and locations accurately [104]. Recognizing objects presented recently in movies selectively activates the HPC when it is necessary to reach back across an event boundary to identify the object [87]. On this view, the HPC is specialized for binding together activated representations instantiated by other brain systems, particularly cortical representations of features of events that could combine in multiple different possible ways. Ranganath and Ritchey [105] have proposed, specifically, that the HPC's interactions with adjacent regions in the medial temporal lobes, and in the medial posterior cortex, are crucial for binding information about individual entities with information about spatiotemporal context to form a coherent event representation. In their account, the perirhinal cortex is the main hub for representing entity information, the parahippocampal cortex and retrosplenial cortex are the main hubs for representing context, and the HPC binds their representations together into an event model. Each of these systems is proposed to interact with a distinct set of lateral cortical regions. Consistent with this view, Baldassano and colleagues [54] found that shifts in cortical patterns in both sets of lateral cortical regions were associated with phasic increases in the HPC. Other models suggest that communication between the HPC and the PFC (specifically, the ventrolateral PFC) may also be important for within-event binding: In a sequential picture memory paradigm, Dubrow and Davachi [106] found that interactions between HPC and ventrolateral PFC predicted within-event sequential memory.

Binding information into event models is, of course, crucial for episodic memory formation because it provides a route for a retrieval cue to call up appropriate information from a related previously-experienced episode, as laid out by context models of episodic memory [107]. One possibility is that the within-event binding function of the HPC (in collaboration with other brain systems) and its episodic memory formation function are complementary, reflecting distinct components of the cascade that occurs at an event boundary. Ben-Yakov and colleagues [108–110] have proposed that phasic activity in the HPC at the ends of events implements a "now print" function, compiling accumulated bound event representations in a way that enables their retrieval after a delay. They found that univariate activity in the HPC was time-locked to the ends of movie clips, and that the magnitude of this activity predicted the likelihood of remembering the just-ended clip (see also [111]. This is consistent with the findings of Dubrow and Davachi [106], who found that phasic responses in the HPC at event boundaries in the sequential picture paradigm predicted subsequent sequential memory.

Together, these results suggest a model in which the HPC plays two related but distinct roles in event model maintenance and updating. During an ongoing event, the HPC may bind disparate concurrently-activated cortical representations into a coherent complex. At an event boundary, the HPC may perform a "now print" function [108–110] or "sharpening" function [105] that stabilized the bound pattern in a way that preserves it for subsequent recall, before abandoning the current binding to make room for binding a new working model (see Figure 4).

Action planning and control

Having described a role for working event models in perception and memory, we turn to their role in action planning and control. We start from a working hypothesis that the structured representations that humans and other animals use to control action are the same event models that they use to understand others' actions [112], and that the predictions we need to make as observers seeking to understand overlap heavily with the predictions we need to make in order to guide action adaptively. This is eminently plausible, especially because in naturalistic behavior observation and comprehension are tightly coupled to action. The idea that complex action control requires structured representation of events has a rich history in artificial intelligence [113] and in psychology [114]. Recent work in computational and experimental neuroscience has begun to link structured representations in perception and in action.

In the context of reinforcement learning, it is possible to give a formal analysis of the costs and benefits of maintaining structured representations of situations. One important contrast is between *model-free* and *model-based* learning [115]. In model-free learning, the system attempts to learn directly the optimal action given all the possible states it experiences. This can lead to optimal behavior if the environment is appropriately structured, but the sequential organization of the behavior has to be determined by the structure of the environment—all the system can learn is the overall value associated with each state. In contrast, a model-based learning system attempts to simultaneously learn the structure of the environment along with the optimal action to be taken in each state. This allows the system to take into account the potential sequence of states the learner might pass through, providing a basis for planning based on predicted future states. The costs and benefits of model-based learning are analogous to the costs and benefits of event models that we described at this paper's opening: Model-based learning is more complex and costlier, but allows the system to learn complicated structures that cannot be tractably learned with model-free learning. Reviewing research on the neural underpinnings of model-based learning, Daw and colleagues [71] associate it with structures in the prefrontal cortex (among others).

One important question about action representations is how the mind and brain capture the hierarchical structure of action. Studies of complex task performance in monkeys and humans suggest that the prefrontal cortex represents the structure of one's current task simultaneously on multiple timescales, possibly with a temporal gradient such that shorter timescales are represented more posteriorly and longer timescales are represented more anteriorly [116–118].

If the same event representations underlie perception and action, then one would expect to find an association between the perception of structure in others' actions and the ability to performed structured actions of one's own. Suggestive evidence comes from recent work on event perception and action control in people with high-functioning autism spectrum disorder (ASD). People with ASD have selective impairments in event segmentation that are not accounted for by general intellectual dysfunction [119]. This population also has a selective impairment in the ability to use a current event context to guide action in a simple decision-making task [120]. However, to our knowledge no one has tested whether individual differences in event perception predict individual differences in action performance in people with ASD.

Bailey and colleagues [25] tested the hypothesis that event perception predicts action performance directly in a population of older adults with varying degrees of cognitive impairment likely due to Alzheimer’s disease (AD). Participants viewed movies of everyday activities and segmented them into events. They also completed a standardized test of naturalistic action performance, which involved packing a child’s lunch and backpack [121]. As expected, healthy participants performed better on both the action perception and the action performance task. However, after controlling for differences in clinical cognitive impairment, those who segmented better were still substantially more successful in the action performance task. A converging result comes from a study of intellectually disabled people in a sheltered workplace, conducted by Sebastian and colleagues [122]. They found that a measure of event segmentation predicted workers’ abilities to perform an assembly task. See Box 5 for a further discussion of the potential utility of event cognition measures for diagnostic purposes and as a target for clinical intervention.

Conclusion

The Oxford English Dictionary [137] gives four senses of the word “parsimony.” One, rarely seen today, has to do with stinginess: “economy of action, effort, or process in an organism or natural system.” Another is the definition familiar to scientists: “the principle that no more entities, causes, or forces than necessary should be invoked in explaining a set of facts or observations.” Event models are unparsimonious in both senses: They are an added burden for an organism to construct, maintain, and update, and they add complexity to scientific accounts of perception, memory, and action. Nonetheless, the evidence drives to the conclusion that they are an important component of human mental life and a valuable theoretical construct for cognitive scientists (see “Outstanding Questions”). The research reviewed in this article provides strong evidence for the existence of event models and provides insight as to the types of problems that these cognitive representations may have evolved to deal with. Not only do event models afford the ability to understand and make predictions about complex systems, including social interactions and relations between actions and objects, but they also allow for the execution of efficient action plans. Although event models are computationally costly, the cost of creating, maintaining, and updating an event model is balanced by the benefit of being able to more easily the solve difficult scenarios that are part and parcel of everyday life.

Acknowledgments

Preparation of this article was supported by the following funding sources: NIH Grant R21 AG041419 and NIH Grant F32 AG050400. We would like to thank Aya Ben-Yakov, Todd Braver, James Cutting, Lisa Feldman Barrett, Katie Hoemann, Michael Kubovy, Joe Magliano, Anna Schapiro, members of the Dynamic Cognition Lab, and two anonymous reviewers for helpful comments on earlier versions of this paper.

References

1. Radvansky, GA., Zacks, JM. Event cognition. Oxford University Press; 2014.

2. Kurby CA, Zacks JM. Segmentation in the perception and memory of events. Trends Cogn. Sci. 2008; 12:72–79. [PubMed: 18178125]

3. Raichle ME. Two views of brain function. Trends Cogn. Sci. 2010; 14:180–190. [PubMed: 20206576]

4. Pavlov, I. Conditioned reflexes. Oxford University Press; 1927.

5. Bar M. The proactive brain: Using analogies and associations to generate predictions. Trends Cogn. Sci. 2007; 11:280–289. [PubMed: 17548232]

6. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 2013; 36:181–204. [PubMed: 23663408]

7. Bar M. The proactive brain: Memory for predictions. Philos. Trans. R. Soc. B Biol. Sci. 2009; 364:1235–1243.

8. Graf M, et al. Predicting point-light actions in real-time. NeuroImage. 2007; 36:T22–T32. [PubMed: 17499167]

9. Shadmehr R, et al. Error correction, sensory prediction, and adaptation in motor control. Annu. Rev. Neurosci. 2010; 33:89–108. [PubMed: 20367317]

10. Henderson JM. Gaze control as prediction. Trends Cogn. Sci. 2017; 21:15–23. [PubMed: 27931846]

11. Pickering M, Garrod S. Do people use language production to make predictions during comprehension? Trends Cogn. Sci. 2007; 11:105–110. [PubMed: 17254833]

12. Maia TV. Reinforcement learning, conditioning, and the brain: Successes and challenges. Cogn. Affect. Behav. Neurosci. 2009; 9:343–364. [PubMed: 19897789]

13. Runeson S, Frykholm G. Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. J. Exp. Psychol. Gen. 1983; 112:585–615.

14. Wolpert DM, Flanagan JR. Motor prediction. Curr. Biol. 2001; 11:R729–R732. [PubMed: 11566114]

15. Jeannerod M. Mental imagery in the motor context. Neuropsychologia. 1995; 33:1419–1432. [PubMed: 8584178]

16. Harnad S. The symbol grounding problem. Phys. Nonlinear Phenom. 1990; 42:335–346.

17. Bransford JD, et al. Sentence memory: A constructive versus interpretive approach. Cognit. Psychol. 1972; 3:193–209.

18. Kintsch W. The role of knowledge in discourse comprehension: A construction-integration model. Psychol. Rev. 1988; 95:163–182. [PubMed: 3375398]

19. Zwaan RA, Radvansky GA. Situation models in language comprehension and memory. Psychol. Bull. 1998; 123:162–185. [PubMed: 9522683]

20. McDaniel MA, et al. What is minimal about predictive inferences? Psychon. Bull. Rev. 2001; 8:840–846. [PubMed: 11848608]

21. O'Brien, EJ., Myers, JL. Text comprehension: A view from the bottom up. In: Goldman, SR., et al., editors. Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso. Erlbaum; 1999. p. 35-53.Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso

22. Johnson-Laird, PN. Mental models: Towards a cognitive science of language, inference, and consciousness. Harvard University Press: 1983.

23. Altmann GT, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition. 1999; 73:247–264. [PubMed: 10585516]

24. Crocker MW, et al. Situated sentence processing: The coordinated interplay account and a neurobehavioral model. Brain Lang. 2010; 112:189–201. [PubMed: 19450874]

25. Bailey, HR., et al. Action perception predicts action performance. Neuropsychologia. 2013. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2013.06.022i

26. Hassabis D, Maguire EA. Deconstructing episodic memory with construction. Trends Cogn. Sci. 2007; 11:299–306. [PubMed: 17548229]

27. Szpunar KK, et al. Contextual processing in episodic future thought. Cereb. Cortex. 2009; 19:1539–1548. [PubMed: 18980949]

28. Intraub, H. Psychology of Learning and Motivation. Vol. 52. Elsevier; 2010. Rethinking scene perception: A multisource model; p. 231-264.
29. Mullally SL, et al. Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. Curr. Biol. 2012; 22:261–268. [PubMed: 22264610]
30. Schacter DL, et al. The future of memory: Remembering, imagining, and the brain. Neuron. 2012; 76:677–694. [PubMed: 23177955]
31. Buckner RL, Carroll DC. Self-projection and the brain. Trends Cogn. Sci. 2007; 11:49–57. [PubMed: 17188554]
32. Barsalou LW. Perceptions of perceptual symbols. Behav. Brain Sci. 1999; 22:637–660.
33. Zwaan RA. Mental simulation in language comprehension and social cognition. Eur. J. Soc. Psychol. 2009; 39:1142–1150.
34. Kemmerer D, et al. Neuroanatomical distribution of five semantic components of verbs: Evidence from fMRI. Brain Lang. 2008; 107:16–43. [PubMed: 17977592]
35. Pulvermüller F. Opinion: Brain mechanisms linking language and action. Nat. Rev. Neurosci. 2005; 6:576–582. [PubMed: 15959465]
36. Willems RM, et al. Neural dissociations between action verb understanding and motor imagery. J. Cogn. Neurosci. 2010; 22:2387–2400. [PubMed: 19925195]
37. Calvo-Merino B. Action observation and acquired motor skills: An fMRI study with expert dancers. Cereb. Cortex. 2004; 15:1243–1249. [PubMed: 15616133]
38. Speer NK, et al. Reading stories activates neural representations of perceptual and motor experiences. Psychol. Sci. 2009; 20:989–999. [PubMed: 19572969]
39. Zacks JM, et al. The brain's cutting-room floor: Segmentation of narrative cinema. Front. Hum. Neurosci. 2010; 4:1–15. [PubMed: 20204154]
40. Mahon BZ, Caramazza A. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J. Physiol.-Paris. 2008; 102:59–70. [PubMed: 18448316]
41. van den Broek P. Integrating memory-based and constructionist processes in accounts of reading comprehension. Discourse Process. 2005; 39:299–316.
42. Zacks JM, et al. Human brain activity time-locked to perceptual event boundaries. Nat. Neurosci. 2001; 4:651–655. [PubMed: 11369948]
43. Newtson D. Attribution and the unit of perception of ongoing behavior. J. Pers. Soc. Psychol. 1973; 28:28–38.
44. Zacks JM, et al. Perceiving, remembering, and communicating structure in events. J. Exp. Psychol. Gen. 2001; 130:29–58. [PubMed: 11293458]
45. Magliano JP, et al. Indexing space and time in film understanding. Appl. Cogn. Psychol. 2001; 15:533–545.
46. Pettijohn KA, Radvansky GA. Narrative event boundaries, reading times, and expectation. Mem. Cognit. 2016; 44:1064–1075.
47. Zacks JM, et al. Using movement and intentions to understand human activity. Cognition. 2009; 112:201–216. [PubMed: 19497569]
48. Zacks JM, et al. Segmentation in reading and film comprehension. J. Exp. Psychol. Gen. 2009; 138:307–327. [PubMed: 19397386]
49. Sargent JQ, et al. Event segmentation ability uniquely predicts event memory. Cognition. 2013; 129:241–255. [PubMed: 23942350]
50. Zacks JM, et al. Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychol. Aging. 2006; 21:466–482. [PubMed: 16953710]
51. Speer NK, et al. Human brain activity time-locked to narrative event boundaries. Psychol. Sci. 2007; 18:449–455. [PubMed: 17576286]
52. Whitney C, et al. Neural correlates of narrative shifts during auditory story comprehension. NeuroImage. 2009; 47:360–366. [PubMed: 19376237]
53. Hasson U, et al. A hierarchy of temporal receptive windows in human cortex. J. Neurosci. 2008; 28:2539–2550. [PubMed: 18322098]
54. Baldassano C, et al. Discovering event structure in continuous narrative perception and memory. bioRxiv. 2016; doi: 10.1101/081018

55. O'Keefe J. Place units in the hippocampus of the freely moving rat. Exp. Neurol. 1976; 51:78–109. [PubMed: 1261644]
56. MacDonald CJ, et al. Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron. 2011; 71:737–749. [PubMed: 21867888]
57. Miller JF, et al. Neural activity in human hippocampal formation reveals the spatial context of retrieved memories. Science. 2013; 342:1111–1114. [PubMed: 24288336]
58. Hsieh L-T, et al. Hippocampal activity patterns carry information about objects in temporal context. Neuron. 2014; 81:1165–1178. [PubMed: 24607234]
59. Raymont V, et al. "Studying injured minds" – The vietnam head injury study and 40 years of brain injury research. Front. Neurol. 2011; 2
60. Zacks JM, et al. Effects of penetrating traumatic brain injury on event segmentation and memory. Cortex. 2016; 74:233–246. [PubMed: 26704077]
61. Kelley HH. Attribution theory in social psychology., presented at the Nebraska symposium on motivation. 1967
62. Kelley HH. The processes of causal attribution. Am. Psychol. 1973; 28:107.
63. Swallow, KM., et al. The effect of viewpoint in event segmentation; Proceedings of the 38th Annual Conference of the Cognitive Science Society; Philadelphia, PA: 2016.
64. Barrett, LF. The theory of constructed emotion: An active inference account of interoception and categorization. Soc. Cogn. Affect. Neurosci. 2016. at <https://www.ncbi.nlm.nih.gov/pubmed/27798257>
65. Hoemann K, et al. Mixed emotions in the predictive brain. Curr. Opin. Behav. Sci. in press.
66. Gendron M, Barrett LF. Emotion perception as conceptual synchrony. Emot. Rev. in press.
67. Hard BM, et al. Making sense of abstract events: building event schemas. Mem. Cognit. 2006; 34:1221–1235.
68. Kurby CA, Zacks JM. Age differences in the perception of hierarchical structure in events. Mem. Cognit. 2011; 39:75–91.
69. Kubovy M. The deep structure of lives. Philos. Sci. 2015; 19:153–176.
70. Zacks JM, et al. Event perception: A mind-brain perspective. Psychol. Bull. 2007; 133:273–293. [PubMed: 17338600]
71. Daw ND, et al. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 2005; 8:1704–1711. [PubMed: 16286932]
72. Heft, H. The ecological approach to navigation: a gibsonian perspective. In: Portugali, J., editor. The construction of cognitive maps. Kluwer Academic Publishers; 1996. p. 105-132.The construction of cognitive maps
73. Newtson, D. Foundations of attribution: The perception of ongoing behavior. In: Harvey, JH., et al., editors. New directions in attribution research. Lawrence Erlbaum Associates; 1976. p. 223-248.
74. Schapiro AC, et al. Neural representations of events arise from temporal community structure. Nat. Neurosci. 2013; 16:486–492. [PubMed: 23416451]
75. Schapiro AC, et al. Statistical learning of temporal community structure in the hippocampus. Hippocampus. 2016; 26:3–8. [PubMed: 26332666]
76. Zacks JM, et al. Prediction error associated with the perceptual segmentation of naturalistic events. J. Cogn. Neurosci. 2011; 23:4057–4066. [PubMed: 21671745]
77. Schultz W, Dickinson A. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 2000; 23:473–500. [PubMed: 10845072]
78. Ezzyat Y, Davachi L. What constitutes an episode in episodic memory? Psychol. Sci. 2011; 22:243–252. [PubMed: 21178116]
79. Speer NK, Zacks JM. Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts•. J. Mem. Lang. 2005; 53:125–140.
80. Morrow DG, et al. Updating situation models during narrative comprehension. J. Mem. Lang. 1989; 28:292–312.
81. Zwaan RA, et al. Constructing multidimensional situation models during reading. Sci. Stud. Read. 1998; 2:199–220.

82. Gernsbacher, MA. Language comprehension as structure building. L. Erlbaum; 1990.

83. Zwaan RA, et al. The construction of situation models in narrative comprehension: An event-indexing model. Psychol. Sci. 1995; 6:292–297.

84. Kurby CA, Zacks JM. Starting from scratch and building brick by brick in comprehension. Mem. Cognit. 2012; 40:812–826.

85. Bailey HR, Zacks JM. Situation model updating in young and older adults: Global versus incremental mechanisms. Psychol. Aging. 2015; 30:232–244. [PubMed: 25938248]

86. Huff M, et al. Changes in situation models modulate processes of event perception in audiovisual narratives. J. Exp. Psychol. Learn. Mem. Cogn. 2014; 40:1377–1388. [PubMed: 24820670]

87. Swallow KM, et al. Changes in events alter how people remember recent information. J. Cogn. Neurosci. 2011; 23:1052–1064. [PubMed: 20521850]

88. Swallow KM, et al. Event boundaries in perception affect memory encoding and updating. J. Exp. Psychol. Gen. 2009; 138:236–257. [PubMed: 19397382]

89. Radvansky GA, Copeland DE. Walking through doorways causes forgetting: Situation models and experienced space. Mem. Cognit. 2006; 34:1150–1156.

90. Radvansky GA, et al. Walking through doorways causes forgetting: Further explorations. Q. J. Exp. Psychol. 2011; doi: 10.1080/17470218.2011.571267

91. Kim G, et al. Pruning of memories by context-based prediction error. Proc. Natl. Acad. Sci. 2014; 111:8997–9002. [PubMed: 24889631]

92. Baddeley A. The episodic buffer: a new component of working memory? Trends Cogn. Sci. 2000; 4:417–423. [PubMed: 11058819]

93. Ericsson KA, Kintsch W. Long-term working memory. Psychol. Rev. 1995; 102:211–245. [PubMed: 7740089]

94. Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. Brain. 1991; 114:727–741. [PubMed: 2043945]

95. D'Esposito M, Postle BR. The dependence of span and delayed-response performance on prefrontal cortex. Neuropsychologia. 1999; 37:1303–1315. [PubMed: 10530730]

96. Funahashi S, et al. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. 1989; 61:331–349. [PubMed: 2918358]

97. Fuster JM, Alexander GE. Neuron activity related to short-term memory. Science. 1971; 173:652–654. [PubMed: 4998337]

98. Müller NG, et al. Contributions of subregions of the prefrontal cortex to working memory: Evidence from brain lesions in humans. J. Cogn. Neurosci. 2002; 14:673–686. [PubMed: 12167253]

99. Davachi L, et al. Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. Proc. Natl. Acad. Sci. 2003; 100:2157–2162. [PubMed: 12578977]

100. Eichenbaum H, et al. The medial temporal lobe and recognition memory. Annu. Rev. Neurosci. 2007; 30:123–152. [PubMed: 17417939]

101. Squire LR. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychol. Rev. 1992; 99:195–231. [PubMed: 1594723]

102. Diana RA, et al. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci. 2007; 11:379–386. [PubMed: 17707683]

103. Rubin, RD., Cohen, NJ. Memory, relational representations, and the long reach of the hippocampus. In: Hannula, DE., Duff, MC., editors. The Hippocampus from Cells to Systems. Springer International Publishing; 2017. p. 337-366.

104. Watson PD, et al. Spatial reconstruction by patients with hippocampal damage is dominated by relational memory errors. Hippocampus. 2013; 23:570–580. [PubMed: 23418096]

105. Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. Nat. Rev. Neurosci. 2012; 13:713–726. [PubMed: 22992647]

106. DuBrow S, Davachi L. Temporal binding within and across events. Neurobiol. Learn. Mem. 2016; 134:107–114. [PubMed: 27422018]

107. Polyn SM, et al. A context maintenance and retrieval model of organizational processes in free recall. Psychol. Rev. 2009; 116:129–156. [PubMed: 19159151]

108. Ben-Yakov A, Dudai Y. Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. J. Neurosci. 2011; 31:9032–9042. [PubMed: 21677186]

109. Ben-Yakov A, et al. Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. J. Exp. Psychol. Gen. 2013; 142:1255. [PubMed: 23815458]

110. Ben-Yakov A, et al. Shifting gears in hippocampus: Temporal dissociation between familiarity and novelty signatures in a single event. J. Neurosci. 2014; 34:12973–12981. [PubMed: 25253846]

111. Hales JB, Brewer JB. The timing of associative memory formation: frontal lobe and anterior medial temporal lobe activity at associative binding predicts memory. J. Neurophysiol. 2011; 105:1454–1463. [PubMed: 21248058]

112. Hommel B, et al. The Theory of Event Coding (TEC): A framework for perception and action planning. Behav. Brain Sci. 2001; 24:849–937. [PubMed: 12239891]

113. Lansky, AL. Proceedings of the Second Annual Conference on Artificial Intelligence Planning Systems. AAAI Press; 1994. Action-based planning; p. 110-115.Proceedings of the Second Annual Conference on Artificial Intelligence Planning Systemspp

114. Miller, GA., et al. Plans and the structure of behavior. Holt; 1960.

115. Daw, ND. Reinforcement learning models of the dopamine system and their behavioral implications. Carnegie Mellon University; 2003. Unpublished doctoral dissertation

116. Badre D. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn. Sci. 2008; 12:193–200. [PubMed: 18403252]

117. Koechlin E, et al. The architecture of cognitive control in the human prefrontal cortex. Science. 2003; 302:1181–1185. [PubMed: 14615530]

118. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 2001; 24:167–202. [PubMed: 11283309]

119. Zalla T, et al. Perceiving goals and actions in individuals with autism spectrum disorders. J. Autism Dev. Disord. 2013; 43:2353–2365. [PubMed: 23420086]

120. Barbalat G, et al. A specific impairment in cognitive control in individuals with high-functioning autism. J. Psychiatr. Res. 2014; 58:26–35. [PubMed: 25106070]

121. Schwartz M, et al. The Naturalistic Action Test: A standardised assessment for everyday action impairment. Neuropsychol. Rehabil. 2002; 12:311–339.

122. Sebastian K, et al. Understanding the individual cognitive potential of persons with intellectual disability in workshops for adapted work: Cognitive potential in workshops. Appl. Cogn. Psychol. 2017; doi: 10.1002/acp.3315

123. Richmond LL, et al. Event Perception: Translations and Applications. J. Appl. Res. Mem. Cogn. 2017; doi: 10.1016/j.jarmac.2016.11.002

124. Giovannetti T. The clinical importance of understanding and improving eventday cognition in older adults. J. Appl. Res. Mem. Cogn. 2017

125. Giovannetti T, et al. Coffee with jelly or unbuttered toast: Commissions and omissions are dissociable aspects of everyday action impairment in Alzheimer’s disease. Neuropsychology. 2008; 22:235–245. [PubMed: 18331166]

126. Flores S, et al. Event segmentation improves event memory up to one month later. J. Exp. Psychol. Learn. Mem. Cogn. in press.

127. Gold DA, et al. Effects of cues to event segmentation on subsequent memory. Cogn. Res. Princ. Implic. 2016; 2

128. Zalla T, et al. Perception of dynamic action in patients with schizophrenia. Psychiatry Res. 2004; 128:39–51. [PubMed: 15450913]

129. Eisenberg ML, et al. Posttraumatic stress and the comprehension of everyday activity. Collabra. 2016; 2:1–10.

130. Zor R, et al. Obsessive-compulsive disorder: a disorder of pessimal (non-functional) motor behavior: OCD: a disorder of pessimal behavior. Acta Psychiatr. Scand. 2009; 120:288–298. [PubMed: 19291081]

131. Sherrill A, Magliano J. Psychopathology applications of event perception basic research: Anticipating the road ahead using Posttraumatic Stress Disorder as an example. J. Appl. Res. Mem. Cogn. 2017

132. Zalla T. Event perception as a building block of social cognition. J. Appl. Res. Mem. Cogn. 2017

133. Khemlani SS, et al. Episodes, events, and models. Front. Hum. Neurosci. 2015; 9

134. Goldin-Meadow S. Studying the mechanisms of language learning by varying the learning environment and the learner. Lang. Cogn. Neurosci. 2015; 30:899–911. [PubMed: 26668813]

135. Alibali MW, Kita S. Gesture highlights perceptually present information for speakers. Gesture. 2010; 10:3–28.

136. Defina R. Events in language and thought: The case of serial verb constructions in Avatime. 2016

137. Oxford English Dictionary parsimony, n. [Online]. Available: http://www.oed.com/view/Entry/138169?redirectedFrom=parsimony

Box 1. Embodiment: What is the representational format of event models?

One important question is: To what extent event models are instantiated by the same modality-laden representations that underlie perception and motor control. One view is that structured mental representations, in order to achieve the computational power characteristic of human thought, need to be abstracted from the modalities of action and perception (Fodor, 1975). An alternative view notes that perceptual and motor systems provide a powerful platform for representing the world, and proposes that for the sake of efficiency cognitive representations re-use these platforms to comprehend, remember, and predict [32, 33].

In language comprehension, neurophysiological studies suggest that when people read about events they activate neural representations of motor properties of those actions [34–36]. In action observation, experts in a motor skill such as dance show greater motor activation compared to novices when watching activities from their domain of expertise, suggesting they are using knowledge to form embodied representations [37]. In our laboratory, we have used fMRI to record brain activity while people read about everyday activities or watch movies, and have been struck by the parallels in the content-specific transient activations associated with features such as interacting with objects or changing locations [38, 39]. These results suggest that, at least some of the time, people do re-use circuits for online action and perception in an offline mode to represent the contents of events for language or visual comprehension.

This interpretation has been challenged, however. Mahon and Caramazza [40] noted that the fact that people tend to activate perceptual and motor representations when thinking about concepts with perceptual and motor contents does not entail that the perceptual and motor representations are causally related to the conceptualization. Instead, conceptual representations could be purely disembodied, but could tend to activate embodied representations through repeated association. Relatedly, one could worry that the task conditions in which such effects have been observed may have task demands that encourage strategic use of perceptual or motor representations in a way that is not characteristic of normal comprehension.

We think that, going forward, it will be critical to test embodied comprehension hypotheses using methods that directly assay representational format, in situations that are as naturalistic as possible. Moreover, we also think it is important to recognize that there may be many comprehension situations in which readers or observers do not construct rich event models [41]—they may skim an instruction manual for a key piece of information, paying little attention to the action sequences described in the text, or search a room for a dropped wallet, paying little attention to the activities going on around them. In these situations, there may be little evidence of event model construction, embodied or otherwise.

An effective working model needs to be relatively stable through the duration of an event —if the model's state is as dynamic as the inputs from which it is constructed, it cannot enable smooth prediction. For example, if one observes someone else buttering a piece of bread, one will benefit from a working model that does not get disrupted by each swipe of the knife. However, to be effective, a working model also needs to be updated when the represented situation transitions from one configuration to another—when the bread is buttered and the diner turns to drink water. Thus, working models need a particular kind of dynamics: stability punctuated by phasic transitions. These transitions segment ongoing experience into a succession of events represented by successive working models.

Box 2. Hierarchies and strands

For the most part, in this paper we describe events in terms of a single segmentation of a timeline into chunks. This reflects the state of the science; however, it is clear that events have structure beyond a single sequence of chunks. One important feature of naturalistic activity is that it can be segmented according to different timescales. Within a single event, segmentation occurs simultaneously on multiple timescales, and different timescales are related hierarchically such that fine-grained events group into larger coarse-grained events [44, 67, 68]. For example, a coarse-grained event within the 'getting ready for work' activity might be 'brushing teeth.' Within this coarse event, fine events might include putting toothpaste on the toothbrush or picking up the toothbrush.

Another important feature of naturalistic activity is that it includes thematically related events that are discontinuous in time. For example, a sculptor may work on a piece in sessions separated by hours, days, and weeks, but conceptualize the sculpting as one unified activity. However, most extant models lack mechanisms to account for structures that cohere despite being discontinuous in time. Kubovy [69] has recently called attention to this problem, and proposed a framework for describing such event structures. Within Kubovy's 'concurrent strands' framework, one might have different representational strands for the many facets of one's life, including home, work/school, errands and personal business.

Strands are specified according to the individual's point of view [69]. Each strand is characterized in three ways: by the individual's role, by the other people typically involved in activities that take place in this strand, and by where activities within this strand typically occur. For example, the 'home' strand may be characterized by the individual's role in the nuclear family, the other individuals that live in the home, and the dwelling that is the setting for many of the 'home' activities. Importantly, Kubovy suggested that one's experience of events within a strand is that events within the same strand are causally related to one another. Even when a time gap is introduced, such as going to work interrupting the home thread, events that occur after the gap are experienced as being related to the events that preceded it in that strand.

Individuals can also transiently activate different strands—being interrupted at work by a reminder to set up a play date for the children ('home' strand) wouldn't be understood by the individual as comprising a causal relation between other events within the 'work' strand and the play date. Instead, the play date would be integrated into the causal chain of events experienced within the 'home' strand even though this event takes place away from the typical setting of events within this strand. In other words, although these events are experienced in a temporally interleaved fashion from the vantage point of the outside observer, the actor tags these events with different meanings and assigns each event to a different strand, thus separating these events experientially. An important question for future research is: How does organization by strands interact with hierarchical organization?

Box 3. Mechanisms of event segmentation

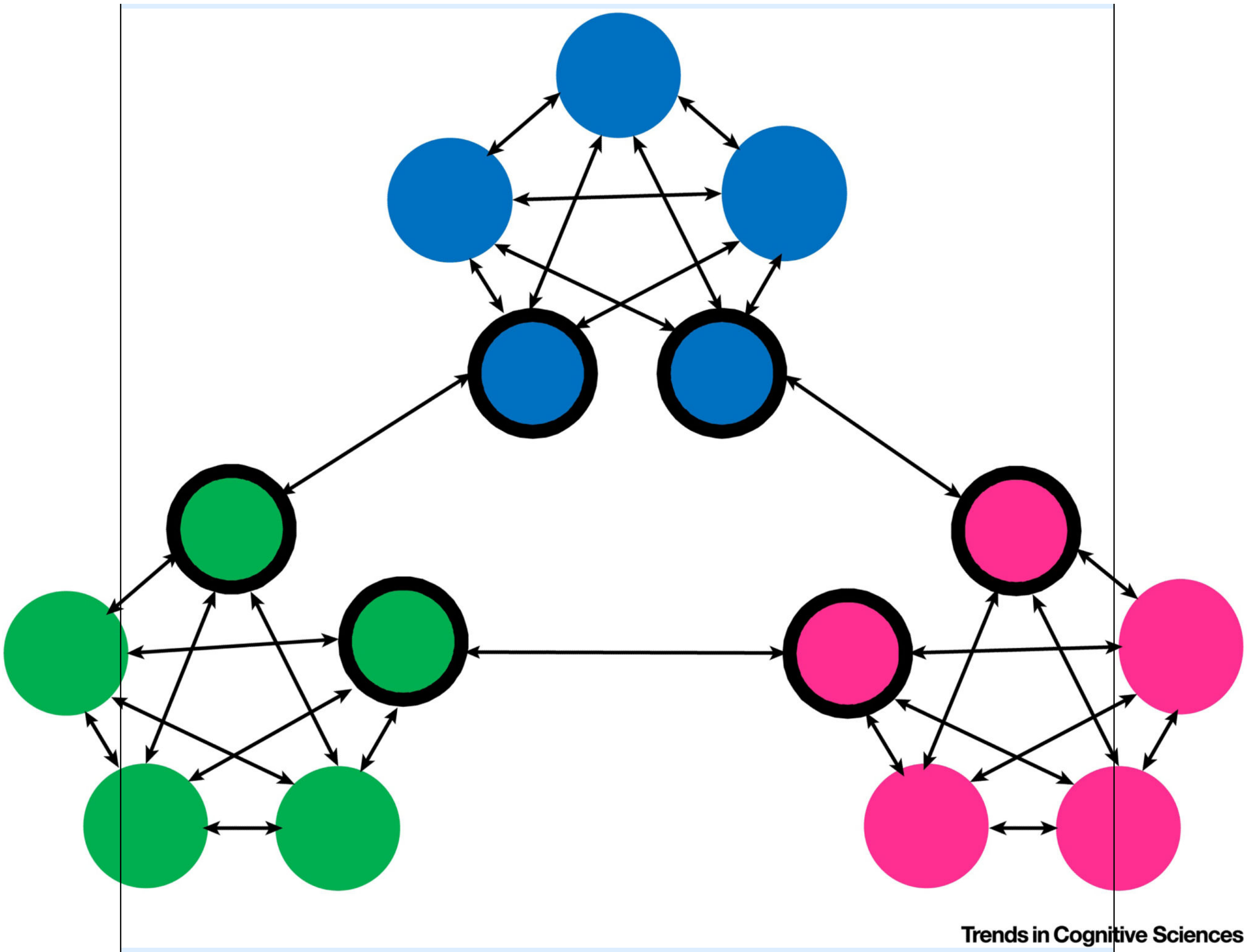
Event segmentation theory [70] proposes a specific mechanism responsible for event segmentation. According to EST, working models are updated in response to spikes in prediction error. However, other mechanisms are possible.

Consider an agent that makes predictions about how activity will unfold, based on some combination of learning from previous experience and inbuilt predispositions. At any point in time, t, the agent makes a prediction about what will happen a short time later, $t+1$. We denote the agent's prediction at time t p_t (where p refers to representation of a state of affairs). We describe the state of affairs the system subsequently experiences as s_{t+1}. The magnitude of prediction error, then, is given by the sum of squared error between p_t and s_{t+1}. (In reinforcement learning models, this quantity is referred to as *surprisal*; [6].) In EST, spikes in this quantity trigger working model updating.

As just stated, p_t is a point estimate. Cognitive systems may not just produce a point prediction but also generate a range of values based on previous experience and the system's dispositions. If that range is large, this means the system has greater *uncertainty* [71]. A second possibility is that this high level of uncertainty about the prediction could trigger the updating processes. Rather than waiting for a prediction error at time $t+1$, the system updates at time t in anticipation that the upcoming situation is unpredictable.

A third possibility is that, rather than updating based on prediction error or prediction uncertainty, the system simply monitors for *change*. It could constantly compare the current state of affairs, s_t, to the recent past, s_{t-1}, and update when the change from $t-1$ to t is large [72, 73].

Finally, a fourth possibility is that the system tracks the history of sequences of states that it experiences, and clumps states based on their co-occurrence over time [74, 75]. A learning system could group states that tend to co-occur nearby in time in many sequences into *communities*, and to segment a sequence when it transitions from a state associated with one community to a state associated with a different community. In a sense, this amounts to segmenting based on change in the internal representation (rather than in the features of the world).



Box 3, Figure I.

Representation of community structure described by Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick [74]. Each node is associated with a particular state, and each edge is associated with possible transitions between states. States within a cluster tend to connect with one another and not with nodes outside of the cluster; however, it is possible to transition from a state in one cluster to a state in another cluster (represented by the dark outlined nodes).

In naturalistic activity, these mechanisms might make very similar hypotheses about how a system will segment. For most learning systems, prediction error is associated with uncertainty. In naturalistic activity, prediction errors tend to be larger when more features of the activity are changing. All of these features are associated with transitions from one community to another. For this reason, teasing apart these potential mechanisms is a major challenge. Nonetheless, we think it is an important one to tackle, because there are times in naturalistic activity when they do pull apart. For example, though larger changes usually lead to larger prediction errors, sometimes the *absence* of change is what is unpredicted. To make this concrete, imagine a child waiting for an ice cream truck. All summer, the truck has turned at the corner, onto the child's street—so entropy is low.

Today the truck continues straight. This *lack* of feature change would be associated with a large prediction error (as well, presumably, a disappointed child).

One piece of evidence for the role of prediction error in event boundaries comes from a series of experiments from our laboratory [76]. In these, participants watched movies that had previously been segmented by other viewers. From time to time, each movie was stopped and participants were asked to predict what would happen in 5 s, by selecting one of two still pictures. EST entails that prediction should be more difficult when participants have to predict across an event boundary, and that is what was found. fMRI indicated that these conditions were associated with greater activity in the *midbrain dopamine system*, which is involved in signaling prediction errors [77].

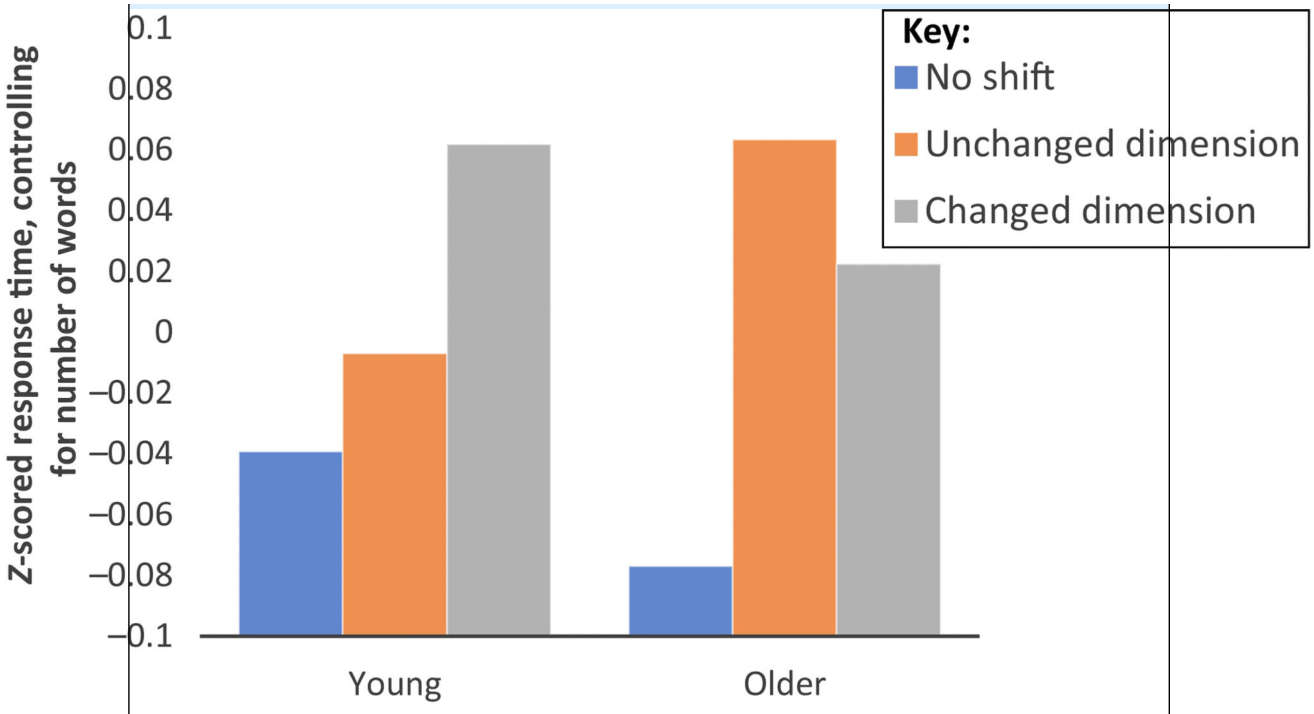
Turning to the updating process, the conditions and mechanisms of working model updating have been a topic of intense focus in recent research. However, they have antecedents in classical studies of discourse comprehension. These studies have been reviewed previously [70], so we will summarize only briefly (see also Box 4 for a discussion of incremental vs global updating). The main conclusion from this work was that in narrative text, when relevant dimensions of the situation change, readers update their working models. EST views these shifts as event boundaries; incoming information is less predictable than it was before, which should lead to working model updating. The most-studied dimensions are time, space and characters. After a shift in a narrative dimension, information presented prior to the shift is often less accessible, as measured by accuracy or response time on recognition tests or by reading speed (.e.g., [78, 79]. However, it is important to mention two important qualifications to this general pattern of results. First, features other than narrative shifts contribute substantially to memory accessibility. Readers often retrieve information that is no longer in a current working model in order to infer information that is missing from the local text, and sometimes simply reading a semantically related phrase is sufficient to retrieve information not in one's working model [21]. Second, shifts do not always impair memory accessibility. For example, there is evidence that readers often do not track spatial shifts, and as a result information may be just as accessible after a spatial shift as before [80, 81]. Which narrative dimensions a reader tracks depends on that reader's reading goals and previous knowledge.

Box 4. Incremental and global working model updating

An important distinction regarding the updating of working models is between *incremental* and *global* updating mechanisms. An incremental updating mechanism replaces information in a working model selectively, affecting only unpredicted or changed information. By contrast, global updating affects all aspects of an working model, meaning that information that may be “innocent” of having caused an update can be affected along with information that led to the update.

Theories of narrative comprehension propose either solely incremental updating, or a mix of incremental and global updating [82, 83]. EST, however, includes only a global updating mechanism [70]. Surprisingly little empirical research has attempted to dissociate incremental from global updating. This is a substantial limitation, because the two mechanisms could contribute differently in different comprehension situations, and could be affected differently by neural dysfunction and individual differences in cognition.

Two recent studies have begun to address this gap using text comprehension paradigms. Kurby and Zacks [84] used a think-aloud paradigm in which participants described their thoughts after reading each phrase in a narrative. They found that when one aspect of the narrated situation changed—for example, moving from an indoor to an outdoor location—readers were more likely to mention that dimension. However, they were also likely to mention also other aspects that had not changed—for example, the characters or objects. This suggests that the readers were updating those other dimensions. Bailey and Zacks [85] used a recognition memory probe design, in which younger and older participants read narratives with changes in two aspects of the situation: spatial location and characters. From time to time, they were probed to recognize a phrase describing one or the other aspect. Response times suggested that both incremental and global updating were taking place: Readers were slowest when probed for an aspect that had just changed, but were also slowed in responding to unchanged dimensions of the situation compared to a no-change condition. This study also illustrates the potential significance of the distinction between incremental and global updating: Older adults were just as slow in responding to probes of an unchanged dimension as to probes of a changed dimension, suggesting that they relied more on global updating.



Trends in Cognitive Sciences

Box 4, Figure I.

Data from Bailey & Zacks, 2015. Mean z-scored response time for younger and older adults, controlling for item-wise differences in number of words. Probes were presented after no shift, after a shift in the unchanged dimension, and after a shift in the changed dimension. Participants were slower to respond after a shift compared to the no-shift condition. Older adults were slowed similarly whether the probe tapped the dimension that had shifted or the dimension that had remained the same, suggesting that they globally updated both dimensions when either changed. In contrast, younger adults are slower for responding to the changed dimension after the shift compared to unchanged information after the shift, suggesting that they performed both local and global updating.

Incremental and global updating may have different consequences for long term memory. Huff, Meitz, and Papenmeier [86] presented readers with movies that included points with changes in 1, 2, 3 or 4 situational dimensions. They found that subsequent recognition memory was better for points with more changes, and interpreted this as evidence that more information had been updated, incrementally. (However, an alternative possibility is that updating was purely global, but occurred with higher probability when more things were changing.)

Thus, the limited empirical data indicate the importance of distinguishing between incremental and global updating, providing evidence for both in the domain of reading comprehension. Going forward, models such as EST will need to address the role of incremental updating.

The first studies to investigate event-based memory updating in dynamic audiovisual events used clips from commercial cinema, which were pre-tested to identify points that were widely judged to be event boundaries [87, 88]. In these experiments, visual recognition of objects was tested by waiting 5 s after a given object left the screen, then presenting the object and a contextually-appropriate foil and asking participants to make a forced-choice recognition judgment. Participants were less able to identify objects when an event boundary had occurred, and when they could identify the objects, retrieval was associated with selective fMRI activation of the hippocampus and parahippocampal cortex, areas associated with long-term episodic memory.

Recent studies from Radvansky and his colleagues have tested the hypothesis that moving from one spatial setting to another leads to working model updating. The *walking-through-doorways* effect shows that people are slower and less accurate in identifying recently-encountered objects after moving from one room to another. The effect holds after controlling for elapsed time, distance, and visibility [89]. It occurs both for objects that are left in the previous room and for objects that are carried along in a backpack but are not visible. It happens both in virtual reality and in real rooms. It does not seem to be highly dependent on task-specific strategies. Finally, it does not seem to simply be a mismatch between the encoding and retrieval context, because after walking through one doorway, returning to the encoding room does not rescue memory [90].

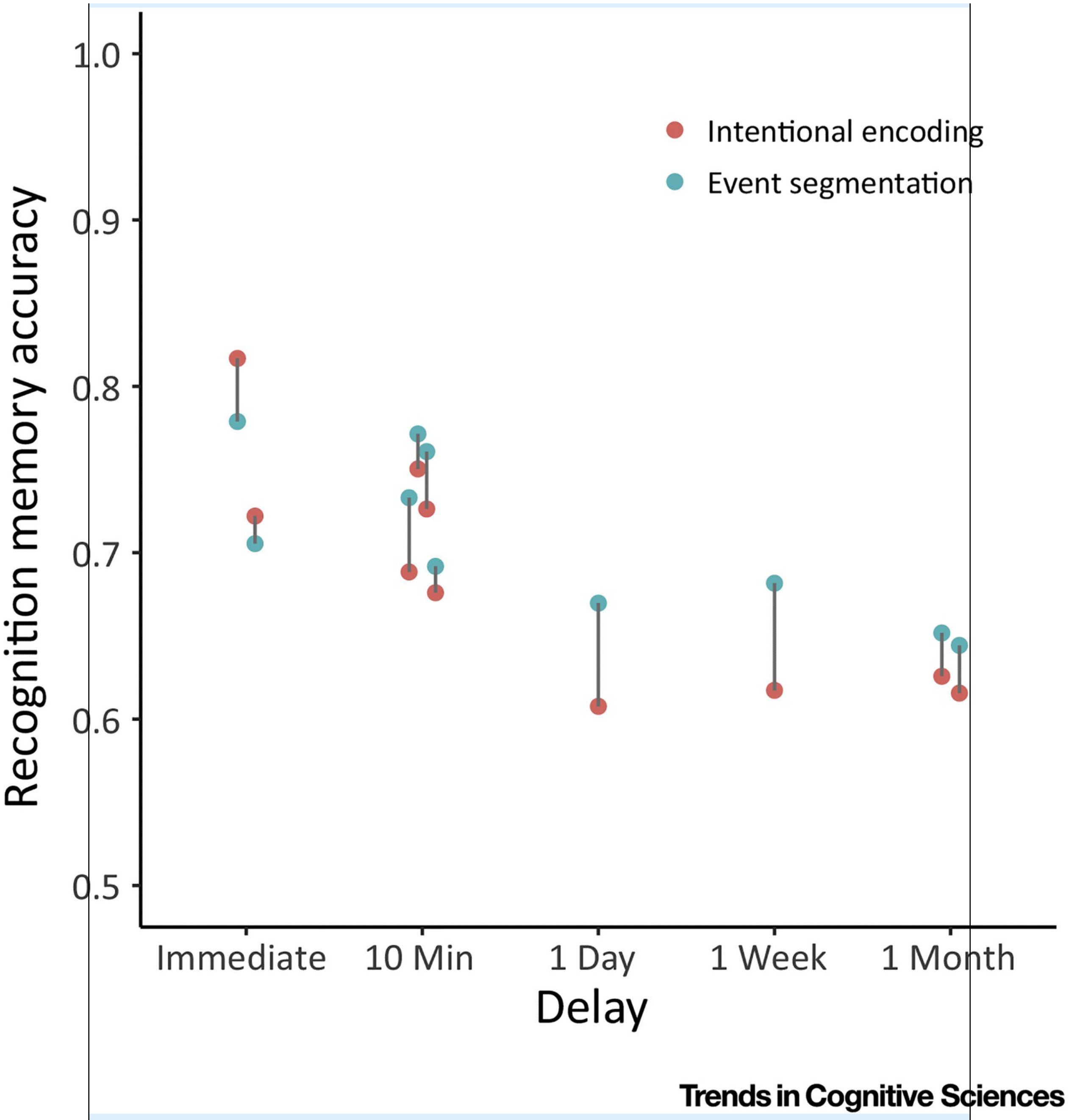
Studies such as these show that working memory representations are updated at event boundaries. However, there is as yet no direct evidence tying this updating to prediction error. One suggestive result comes from a study of long term memory editing [91], in which predictions in the medial temporal lobes were estimated using a multivariate fMRI method. When predictions were strong and wrong, indicating large prediction errors, participants had better delayed memory for the unpredicted information. However, updating effects in working memory have not yet been directly studied.

Thus, spikes in prediction error are thought to be related to working model updating, and when an working model is updated the observer has the phenomenological experience of one meaningful unit of activity coming to an end (and the next unit of activity beginning). Because prediction error is high at event boundaries, people have more difficulty predicting short-term future happenings across event boundaries compared to within an event. When working models are updated, information contained the previous working model becomes less accessible. What are the mechanisms that bind information together within an event, and that allow access to information from a previous event once it has ended? We turn to recent research on these questions now.

Box 5. Diagnostics and intervention

Event models might at first blush seem too “messy” a basis for clinical diagnosis and intervention, but in fact they are quite easy to assess in the clinic and in the laboratory, and could be important diagnostic indicators of dysfunction in everyday life [123]. People differ reliably in their ability to segment everyday events in a normative fashion, and in their ability to subsequently remember those events when tested with recall or recognition measures (e.g., [49]. Event segmentation measures are less influenced by education and SES than many standard neuropsychological tests [124]. It is also possible to measure one’s ability to perform everyday actions reliably, using measures that give a detailed picture of the kinds of action errors a given individual makes [121]. These measures are sensitive to group differences including age, clinical dementia status, and genetic markers of dementia risk [25, 125]. Moreover, event segmentation measures reliably predict both subsequent memory [49] and action performance [25]. Thus, measures of event cognition may be attractive bases for future diagnostic instruments.

Event cognition is also an attractive basis for cognitive interventions. Simply asking individuals to attend to the structure of an event can improve memory for the event. A recent study by Flores and colleagues asked participants either to segment the events depicted in a video, to press a button every few seconds while the video was playing, or simply to intentionally encode the information in the video while passively watching. Memory for the event immediately after viewing did not differ by condition, but for participants in the segmentation condition, memory after a 10-minute delay was significantly better than the other two conditions. This benefit of segmentation for memory persisted through a 1-month delay [126]. Similarly, cueing of event boundaries improved memory for boundary information in both younger and older adults [127].



Box 5, Figure I.

Summary of 5 experiments comparing recognition memory after intentional encoding to memory after event segmentation. When tested immediately, intentional encoding was superior; however, at all other delays event segmentation led to better memory.

If event cognition interventions improve memory in healthy people, they may also prove valuable in the clinic. Deficits in event cognition have been identified in individuals with characteristics of autism [119], schizophrenia [128], Alzheimer's disease [50], post-traumatic stress disorder [129], and obsessive-compulsive disorder [130]. Researchers studying several of these disorders have suggested that event-comprehension interventions have promise in remediating the disorders' symptoms or compensating for their effects [124, 131, 132].

The dual role we are proposing for event models becomes very clear when one considers situations in which a person needs to coordinate with others to accomplish a task. Khemlani, Harrison, and Trafton [133] described an architecture to enable a robot to segment and represent events while coordinating its actions with a team of humans. In this architecture, segmentation is organized by goals, locations, and characters and objects; changes in these features correspond to event boundaries. The system actively maintains a structured representation of the events taking place, which includes information about the physical situation, the robot's goals, and the goals of its human colleagues. In order for the robot to collaborate effectively, its representation needs to correspond with those of its human partners.

To build such shared event representations, people rely heavily on communication, both verbal and nonverbal. Recent work on gesture has shown how it can shape event representations adaptively and efficiently, complementing the affordances of verbal language [134]. People use gesture to highlight information to which we want their conversational partners to attend [135] and to communicate about abstract concepts [134]. The timing of gestures can highlight the segmentation structure of an event description; for example, in languages that can use serial verbs to describe a complex event with a single phrase, gestures cover just those constructions that correspond with single conceptual events [136].

In short, recent studies of action perception and action performance in people with ASD, healthy aging and early AD, and intellectual disability all suggest that individual differences in the ability to form event representations in perception are predictive of individual differences in the ability to perform structured action sequences. Such representations are necessary both for coordinating one's own solitary behavior, and for coordinating behavior with others. The ability to infer goals and mental states and the ability to predict likely actions of others facilitates social coordination. By proactively inferring others' actions on an ongoing basis, the brain places one in a position to act or react appropriately at any moment given the opportunity. In one's own day-to-day life, such representations facilitate the ability to track and plan and carry out complex, goal-directed activities while remaining flexible enough to respond to, and accommodate, changes in the environment that necessitate modification of the action plan.

Trends

Advanced neuroimaging methods and naturalistic stimuli are being used to characterize event representations in extended activities.

Behavioral studies are beginning to characterize event segmentation in interactive, first-person experiences.

Behavioral and neuroimaging studies are characterizing the role of event model updating in working memory access.

Studies of special populations and individual differences are characterizing how event models develop over the lifespan, vary across individuals, and are impaired by disease and injury.

Neuroimaging studies are beginning to characterize interactions between the hippocampus, subcortical structures, and the cortex in binding features into events.

Outstanding Questions

What computational features trigger the updating of working event models, and how is the updating implemented computationally?

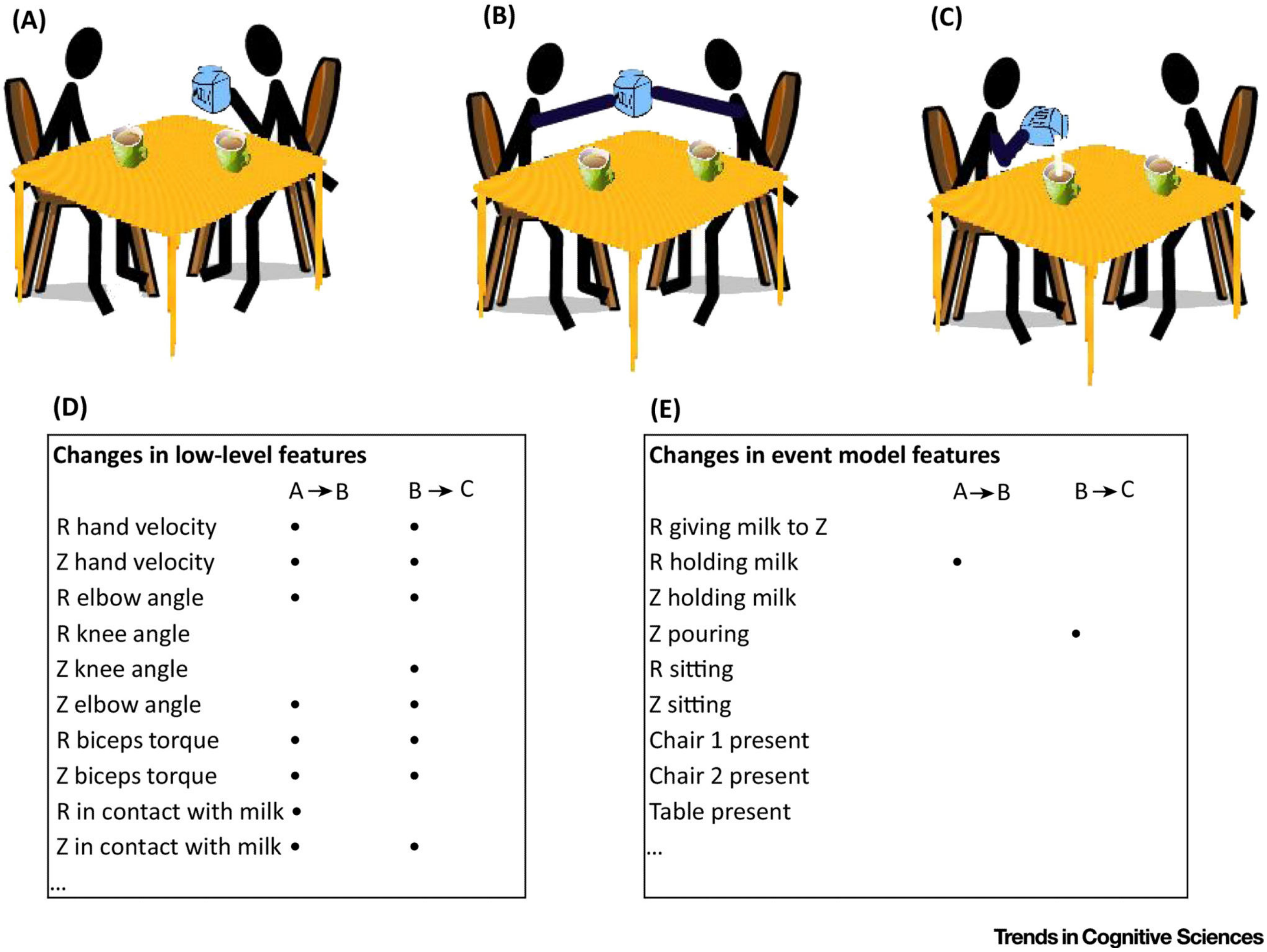
Which aspects of event representations are shared across perception and action control, and which are unique to each domain?

How do event models from real life differ from event models from mediated life, especially movies?

How does retrieval from episodic memory and semantic knowledge affect the construction of working event models?

How are different event strands represented in the mind and brain?

How can event cognition be improved with training or cognitive aids?



Changes in low-level features

	A→B	B→C
R hand velocity	•	•
Z hand velocity	•	•
R elbow angle	•	•
R knee angle		
Z knee angle		•
Z elbow angle	•	•
R biceps torque	•	•
Z biceps torque	•	•
R in contact with milk	•	
Z in contact with milk	•	•
…		

Changes in event model features

	A→B	B→C
R giving milk to Z		
R holding milk	•	
Z holding milk		
Z pouring		•
R sitting		
Z sitting		
Chair 1 present		
Chair 2 present		
Table present		
…		

Trends in Cognitive Sciences

Figure 1.

a) Rebecca (R) is in the middle of passing milk to Zach (Z). b) Z has accepted milk. c) Z is pouring milk. d) Conceptual description of changes in low-level features. Examples include a few different reference frames—body part velocities, joint angles, muscle torques, contact relations—to illustrate that the point is general. e) Conceptual depiction of changes in features of the situation, as might be represented in event models; again, a few representative examples are given. When cast in terms of event model variables, change is smoother and thus more learnable. For example, the sequence "get milk, then pour milk" is easy for many systems to learn, whereas the sequence of velocities or joint angles would be highly challenging.

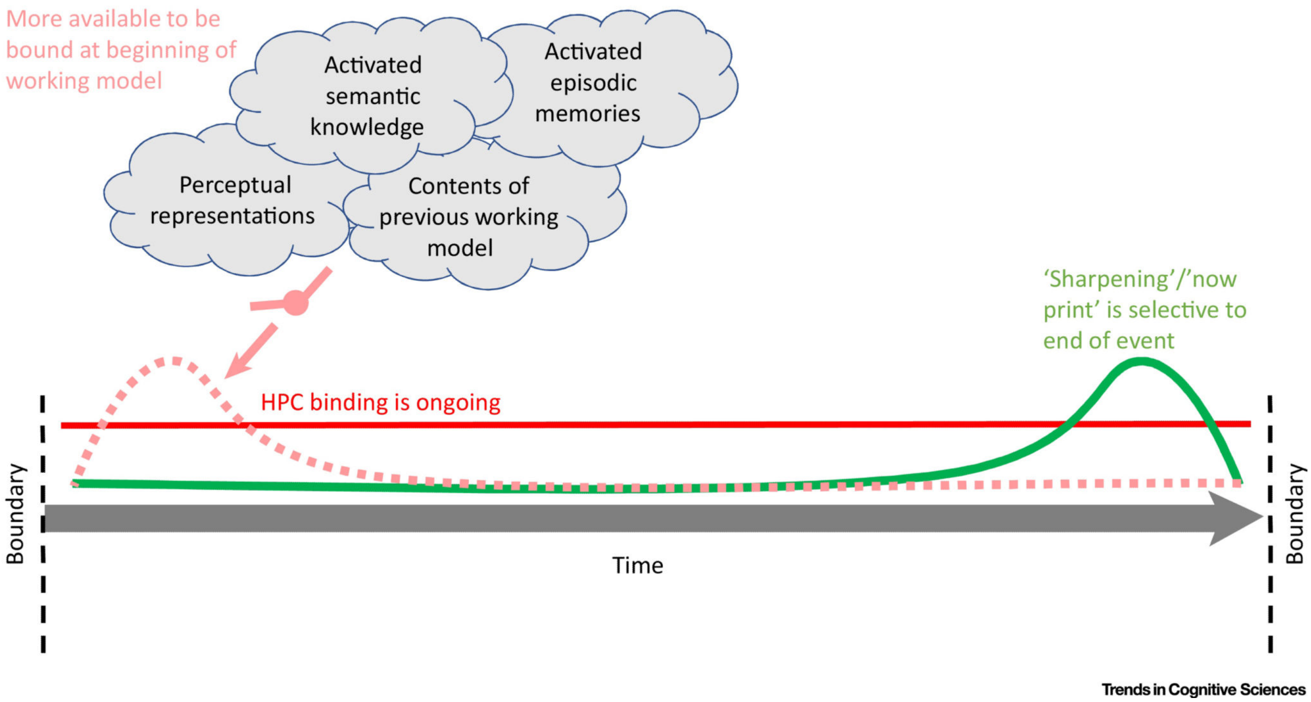


Figure 2.

Two roles for the hippocampus in working models. 1) The HPC binds active cortical representations together. Binding may be ongoing (red), but the information available to be bound disproportionally represents the beginnings of the events due to gating (pink). 2) The “sharpening/now print” function (green) is hypothesized to be selective to the end of the event.