12-15-2019

Prediction-Based Learning and Processing of Event Knowledge.

Ken McRae

Kevin S Brown

Jeffrey L Elman

TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science (2019) 1–18

ISSN:1756-8765 online
DOI: 10.1111/tops.12482

This article is part of the topic "Event-Predictive Cognition: From Sensorimotor via Conceptual to Language-Based Structures and Processes," Martin V. Butz, David Bilkey, and Alistair Knott (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Prediction-Based Learning and Processing of Event Knowledge

Ken McRae,[a] Kevin S. Brown,[b] Jeffrey L. Elman[c]

[a]*Department of Psychology, Brain & Mind Institute, University of Western Ontario*
[b]*Department of Pharmaceutical Sciences, Chemical, Biological, & Environmental Engineering, Oregon State University*
[c]*Department of Cognitive Science, University of California San Diego*

Received 19 May 2019; received in revised form 23 October 2019; accepted 30 October 2019

Abstract

Knowledge of common events is central to many aspects of cognition. Intuitively, it seems as though events are linear chains of the activities of which they are comprised. In line with this intuition, a number of theories of the temporal structure of event knowledge have posited mental representations (data structures) consisting of linear chains of activities. Competing theories focus on the hierarchical nature of event knowledge, with representations comprising ordered scenes, and chains of activities within those scenes. We present evidence that the temporal structure of events typically is not well-defined, but it is much richer and more variable both within and across events than has usually been assumed. We also present evidence that prediction-based neural network models can learn these rich and variable event structures and produce behaviors that reflect human performance. We conclude that knowledge of the temporal structure of events in the human mind emerges as a consequence of prediction-based learning.

Keywords: Event knowledge; Prediction; Connectionist modeling; Network science

Correspondence should be sent to Ken McRae, Department of Psychology and Brain & Mind Institute, University of Western Ontario, Western Interdisciplinary Research Building, London, Ontario, Canada N6A 3K7. E-mail: kenm@uwo.ca

1. Introduction

Theories of cognition have long sought to account for the ways in which human knowledge is organized. Psychological constructs such as *concept* seem well suited for many domains of knowledge. For example, the notion of *concept* is central to most theories of people's knowledge of entities such as cows, objects such as shoes, abstract concepts such as attitude, and ad hoc categories such as *things to take on a camping trip* (Barsalou, 1983; Murphy, 2002). For these domains, taxonomic, similarity, and thematic relations have been central to theories of knowledge organization and class inclusion.

However, humans also develop knowledge about regularities involving behavior and events in the world. How do we go about doing things in our everyday lives? How do we interpret the behavior of others? How do we anticipate what is likely to happen next? When something happens, how do we interpret whether it is coincidental or it reflects cause–effect relationships? Being able to answer such questions allows us to anticipate the consequences of our own actions and those of others, and thus allows us to make inferences about the possible goals or intentions that underlie those actions.

The question of how the temporal dimension of events is represented in human memory has been an important subject of study for quite some time. Perhaps the first detailed computational model of event knowledge was Minsky's (1974) proposal of *frames*. Minsky defined a frame as.

> a data-structure for representing a stereotyped situation . . . attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. (p. 1)

The very explicitness of frame-based models that made them attractive also exposed their Achilles' heel. It became apparent that frames suffered from significant limitations. Most troubling was their brittleness and inflexibility. The restaurant script described a canonical event, but this canonical event rarely if ever occurs because there are a huge number of context-dependent variations. This challenge was recognized from the start (Schank & Abelson, 1977), but the mechanisms that were developed to deal with the variation (e.g., Memory Organization Packets; Schank, 1980) seemed not only *post hoc*, but also suffered from the same brittleness that created their need in the first place.

Furthermore, none of these proposals addressed the question of how event knowledge might be learned in the first place, or how and when event knowledge should be modified as a result of experience (the Piagetian puzzle of when to *accommodate* and when to *assimilate*). In our view, these and related problems reflect important intrinsic limitations of the symbolic, digital architectures that had been used to implement frames and scripts.

Although efforts to develop mechanistic accounts of event models did not entirely cease, work in this area and appeals to schemas and event knowledge declined over time.

However, in the past number of years, event cognition has had a major resurgence. Part of this effort has involved designing and implementing models that learn and use knowledge about events in the service of segmenting the perceptual stream into events (Reynolds, Zacks, & Braver, 2007), learning and processing information about the components of events and their temporal structure (Elman & McRae, 2019), and understanding language (Frank, Koppen, Noordman, & Vonk, 2003; Mayberry, Crocker, & Knoeferle, 2009; Modi, 2016; Venhuizen et al., 2019). A central unifying aspect is that many of these models are based on computational systems that feature prediction as pivotal to both learning and processing. This event-predictive cognition approach demonstrates promise for overcoming important limitations that were inherent to models based on symbolic data structures, and this approach has provided new and nuanced insights into how event knowledge may be learned and used (Butz, Bilkey, Humaidan, Knott, & Otte, 2019).

In Cognitive Science, a number of terms have been used to describe knowledge of common events such as going to a restaurant. These include schema (Anderson, 1978; Norman & Rumelhart, 1981; Rumelhart, 1980), frame (Minsky, 1974), script (Abelson, 1981; Schank & Abelson, 1977), story (Mandler, 1984), and the related notion of situation model (Zwaan & Radvansky, 1998), among others. All of these models of human knowledge are based on symbolic data structures that provide efficient abstractions over many instances, or facilitate construction of new instances on the fly. In this article, we use the term *event knowledge* to denote people's knowledge of events and situations, recognizing that there is both considerable overlap in the types of knowledge captured by these many terms, and multiple ways in which the various models associated with these terms focus on different aspects of event knowledge.

It is important to clarify our terminology up front. Terminology in research on events is somewhat confusing because the notion of "event" plays an important role in numerous areas of Cognitive Science (event memory in cognitive psychology, linguistics, psycholinguistics, action and motor planning, and robotics, to name some of them). That is, the goals of the research and researchers that feature "events" differ greatly across, for example, linguistics, cognitive psychology, and computational models of action planning. This creates a situation in which the "levels" of events, actions, and motor plans, as well as how those levels are labeled, differ across areas and researchers. A researcher's focus on specific levels depends to a great extent on precisely what phenomena she is trying to explain. Various researchers have used terms like event, activity, and action in different ways. In this article, we use *activity* to refer to somewhat abstracted knowledge of what roughly corresponds to an action, such as *mix the ingredients with a spoon*, and we use *event* to refer to a series of activities, such as *baking an apple pie*. We recognize that events, activities, and actions are grounded in sensorimotor experiences, and that understanding this grounding is an important aspect of research on events in Cognitive Science. However, our article focuses on knowledge that resides at a somewhat more abstract level.

There are at least two major dimensions of activities and events. The first is the set of components that make up an activity. Important components include agents, patients, recipients, instruments, and contexts (which often correspond to locations). The second is the temporal order in which a set of activities within an event unfolds. In this article, we focus mainly on people's knowledge of the temporal order of activities; that is, we emphasize time.

There has been a great deal of research into how events are represented in the mind, and this work goes back many years in the history of cognitive psychology (Bower, Black, & Turner, 1979; Minsky, 1974; Schank & Abelson, 1977). As such, a number of theories have been advanced regarding how event knowledge, including the time-course of events, is represented in the mind. As pointed out by Elman (1995), all of our behaviors unfold over time and time is the context within which we understand the world. For example, a primary source of information that allows people to recognize causality concerns the fact that causes typically precede effects. In fact, it is difficult to think about phenomena such as language, action, goal-directed behavior, social behavior, or planning without some way of representing time. A number of highly influential and important theories of event knowledge dealt with time by creating prespecified template-like data structures that were used to represent a sequence of activities (Abelson, 1981). That is, time was encoded representationally. Models such as those based on scripts were relatively brittle and inflexible, and they did not address how event knowledge is learned, although these limitations were acknowledged at the time. On the other hand, Elman (1990) argued for a different approach in which time is represented implicitly. He suggested that time should be represented by the effect that it has on *processing*, rather than being encoded as links in memory, or as part of the input to a model (e.g., a spatial representation of time, as in a shift register). This entails using a model that has temporally dynamic properties that are responsive to temporal sequences. In Elman's simple recurrent network (SRN), the computational units (hidden) at time t serve as part of the input state at time $t + 1$. In addition, his model was trained to predict what might come next given the current input. Thus, processing is influenced both by the current input and the state of the system that resulted from the previous string of inputs. In other words, the system instantiates a temporally sensitive memory that allows it to encode the temporal properties of sequential input because the internal representations are influenced by temporal context, with the effect of time being implicit in the model's internal states.

The main goal of this article is to argue that knowledge of the temporal structure of events in the human mind emerges as a consequence of prediction-based learning and processing. In fact, a number of recent theories that focus on event segmentation, event knowledge, and event-based language comprehension have been implemented using models that process through time and/or have temporally sensitive memory (Botvinick & Plaut, 2004; Elman & McRae, 2019; Modi, 2016; Reynolds, Zacks, & Braver, 2007; Takac & Knott, 2016a, 2016b; Venhuizen, Crocker, & Brouwer, 2019). We discuss these models, with a particular focus on Elman and McRae's connectionist attractor model of event knowledge.

1.1. The temporal nature of events

Intuitively, it feels as though activities that comprise events follow a coherent, regular, and consistent temporal order. For example, for the event *taking money out of an ATM*, you go to the ATM, open your purse or wallet, take out your bank card, insert your bank card into the machine, type in your personal identification number, and so on until you put away your bank card and cash. Because of the strength of this intuition about ordering, the theoretical idea that event representations are composed of a consistent linear sequence of activities has a strong appeal. In fact, it has led to theories in which event representations correspond to linked linear chains of activity nodes in the mind (Barsalou & Sewell, 1985; Drummer, van der Meer, & Schaadt, 2016).

However, there are difficult challenges in understanding the ways in which an event's temporal organization plays a role in its representation and access. The event cognition literature has not been free of empirical controversy, in the sense that divergent findings have been reported regarding both the form and use of event knowledge. One major controversy hinges on the extent to which the temporal structure of events is encoded in long-term memory. This might seem like an odd thing to question, given the recurring theme of events as sequentially structured activity sequences. Indeed, there is evidence for linear chain-like temporal representations of events in memory (Barsalou & Sewell, 1985; Bower, Black, & Turner, 1979; Coll-Florit & Gennari, 2011; Drummer, van der Meer, & Schaadt, 2016; Lancaster & Barsalou, 1997; Raisig, Welke, Hagendorf, & van der Meer, 2007; van der Meer, Beyer, Heinze, & Badel, 2002; Zwaan, 1996). Furthermore, the literature on linguistic aspect reveals a fine-grained sensitivity to the temporal contour of events (Becker, Ferretti, & Madden-Lombardi, 2013; Brennan & Pylkkänen, 2008; Paczynski, Jackendoff, & Kuperberg, 2014; Piñango, Zurif, & Jackendoff, 1999; Todorova, Straub, Badecker, & Frank, 2000).

The notion that an event's temporal organization is directly mirrored in its representation and access has been challenged by studies that failed to find evidence that the temporal structure of events is encoded in a strict chain-like manner in long-term memory (Galambos & Rips, 1982). These data come from experiments in which participants make judgments about an event's activities. They test the hypothesis that if sequential order is the dominant (sole) organizing principle in the representation of event structure, performance should be facilitated in specific ways. For example, Galambos and Rips (1982) compared predictions of a model in which events are represented as linear chains of activities versus a hierarchical model. They found no reliable evidence for linear chains of activities, and concluded that events are not represented solely as linear sequences. Galambos and Rips did, however, find that centrality (a measure of the importance of an activity) facilitated access to activities within an event. This is consistent with evidence of hierarchical encoding of event knowledge that has been found in other studies (Black & Bower, 1980; Bower et al., 1979).

In hindsight, these contradictory findings concerning people's sensitivity to the temporal structure of events may not be that surprising. Although a strong intuition exists that events are comprised of consistent linear sequences, it may be the case that with respect to temporal order, there are multiple sources of variability. For example, there may be a great deal of variability in the regularity of the temporal order of activities across events (this was recognized some time ago when a distinction was made between strong and weak scripts; Abelson, 1981). Some events, like *taking money out of an ATM* or *changing a flat tire*, may be quite consistent and linear. This is related to the fact that one must, for example, take your card out of your wallet prior to inserting it into the ATM, and you must perform both of these actions prior to typing your PIN. On the other hand, for an event like *cleaning the house*, there are many activities that frequently are part of the event such as vacuuming and cleaning the toilets, but in general, the activities that comprise cleaning the house can be completed in many orders, and in the end, the house will be clean.

Single instances of real-world events do indeed live in time, and therefore follow a linear temporal order. However, an event is often interrupted, being intertwined with other activities and events. For example, consider the set of activities that might comprise *making pasta for dinner*. You get the pasta from the cupboard, get a pot from a kitchen drawer, put water on the stove to boil, read and answer a text, get the pasta sauce from the fridge, change the music on your wireless home stereo system, put the pasta in a pot, break up an argument between your 3- and 5-year-old children, set the table, quickly feed the dog, and so on. In other words, as people learn about events that are directly experienced, the input often is noisy, fragmented, and variable.

In addition, much of people's knowledge about events comes from linguistic descriptions of them. Indeed, for learning about many types of events, hearing or reading about them is the primary source of input because we rarely if ever directly experience them. Furthermore, linguistic descriptions often play with time. It is well known that spoken and written descriptions of events do not tend to mirror or respect real time, and language contains a number of vehicles to signal and warp time. Thus, linguistic descriptions of events often provide partial, fragmented, and temporally disjointed information about events. Overall, people's event knowledge is learned from a huge number of highly variable directly experienced and linguistically described examples.

2. Analyzing event knowledge using graph theory

K. S. Brown, N. Christidis, J. L. Elman, and K. McRae (in preparation) investigated event structure by representing events as graphs. They conducted a norming study in which participants produced an ordered set of up to 12 activities for 81 events such as *taking money out of an ATM* and *cleaning the house* (for similar event norms, see Raisig, Welke, Hagendorf, & van der Meer, 2009; Rosen et al., 2003) Participants typed their responses, and the study was conducted using Qualtrics with Master's participants from Amazon Mechanical Turk. K. S. Brown et al. (in preparation) purposely chose events that intuitively differed in their temporal extent and the consistency of their temporal structure. Approximately 25 participants generated activities for each event. The data were collated to standardize the written descriptions of activities that referred to the same activity but were worded slightly differently (e.g., get bank card from wallet and take out bank card). For some events, there was a high degree of consensus regarding the individual activities produced and the ordering of sets of activities.

For other events, however, participants' responses showed a great deal of variability both in the activities they provided, and the order in which they provided them. No two participants generated identical sets of activities for any of the events. This variability most likely results from three sources: (a) variability in the ways in which events unfold; (b) variability in what people attend to, and find important, in an event (event construal); and (c) variability in what participants believe should be reported in any given situation, which can be influenced by what they believe to be mundane or default (and perhaps not reported), or the perspective that they take when listing activities in terms of providing detail or the lack of it. We can see evidence for all of these in our data. As to (a), it is clear that some events can be conducted or construed in multiple ways; participants do not generally agree (nor should we expect them to) on whether you need to eat before versus after you play frisbee at a picnic. An extreme example of (b) comes from *changing a flat tire*, in which one participant's activity sequence consisted solely of "call AAA". This same event, along with *writing an email*, shows signs of variability due to (c) as well, in which some participants listed activities such as "acquire computer," whereas most others assumed the presence of one. Despite the presence of language-based sources of variability, our conclusions that events may be naturally completed in various ways, and that events exist in the mind as construals, are grounded by the data and supported by the modeling.

From participants' generated lists of activities, K. S. Brown et al. (in preparation) created a directed, weighted graph for each event. The graphs were constructed by inputting all sequential pairs of activities in the order in which each individual participant produced them. Therefore, each graph is an amalgamation of the approximately 25 participants' activity lists. The resulting graphs clearly demonstrate that the temporal structure of events is much richer than has typically been assumed. That is, in addition to there being a great deal of variability across the 81 events, there is substantial variability in temporal structure within each event.

Figs. 1 and 2 present the graphs for *writing an email* and *shopping for clothes*. Activities (nodes) are depicted as labeled ovals, and an arrow (directed edge) connects each ordered pair of activities. The darkness and thickness of an arrow from activity A to activity B is proportional to the number of participants that produced an activity list in which A directly preceded B. *Writing an email* (Fig. 1) is a nice example because it contains the least number of unique activities of the 81 normed events (39, making it the easiest to visualize) and it is an event that intuitively seems that it should be well-defined in terms of sequential order. The sequence of dark thick arrows shows the generalized trajectory through this event. On the other hand, even that path is not deterministic. Overall, the variability in how *writing an email* is conducted in the world and/or construed in the mind is readily apparent in that it does not correspond to a linear chain of activities.

Graph theoretical analyses also may provide empirically based insights into hierarchical theories in which events are divided into scenes (Ghosh & Gilboa, 2014; Zacks &

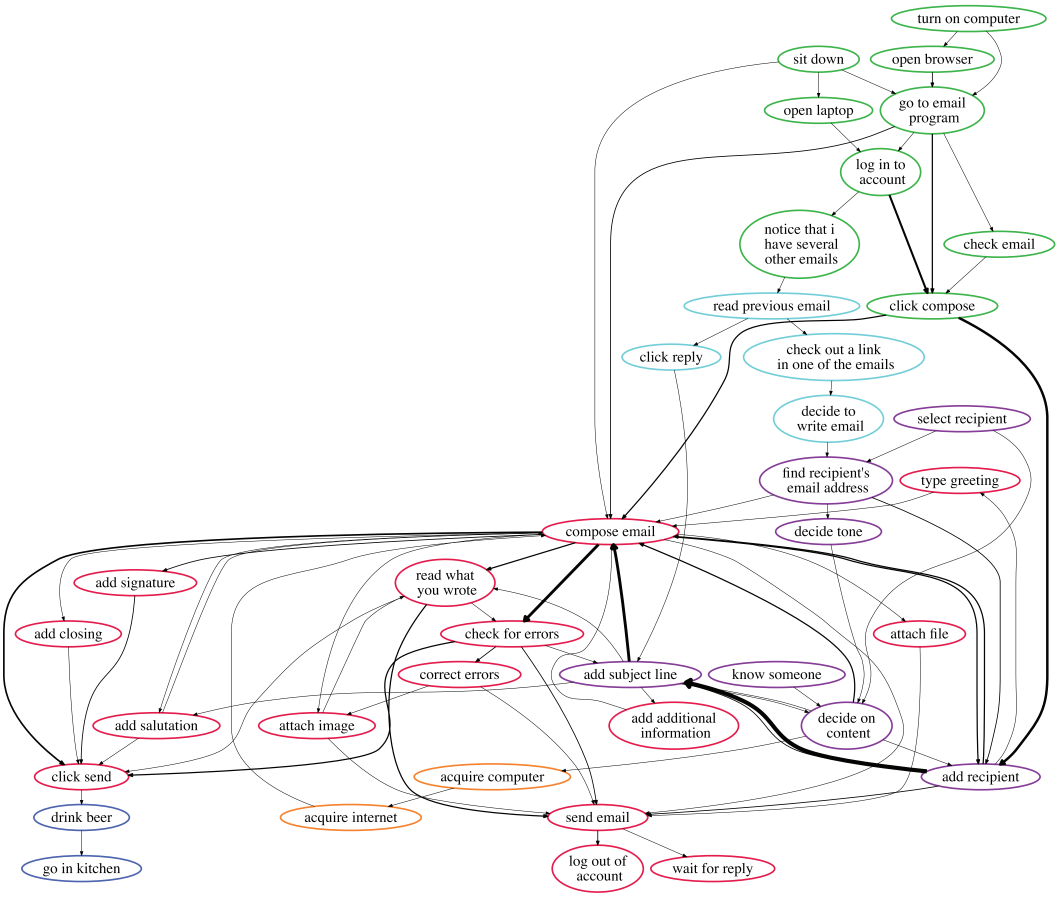

Fig. 1. Graph constructed from participants' produced activities for the event of *writing an email*.

Tversky, 2001). For example, one might imagine that the event *changing a flat tire* may be subdivided into four major hierarchically structured scenes (preparation, removing the flat tire, installing the new tire, finishing up, and driving away). In Figs. 1 and 2, we have labeled the nodes using an algorithm that maximizes graph modularity. Highly modular graphs are "clumpy" and consist of groups of nodes more highly connected to each other than to other groups (Newman & Girvan, 2004). Decomposing a graph into modules or communities is analogous to clustering high dimensional data. In directed weighted graphs, communities consist of sets of nodes that are more strongly directionally connected than is expected by chance (Leicht & Newman, 2008). Fig. 1 (*writing an email*) shows four primary communities/scenes. Green ovals designate activities that correspond to the preparatory phase of writing an email. Light blue and purple ovals specify additional initial phases. The activities in red are the guts of writing an email. Amusingly, there are the activities in darker blue ovals; one participant apparently thought that writing an email requires celebrating the small victories in life by drinking a beer and going into the kitchen (in that order).

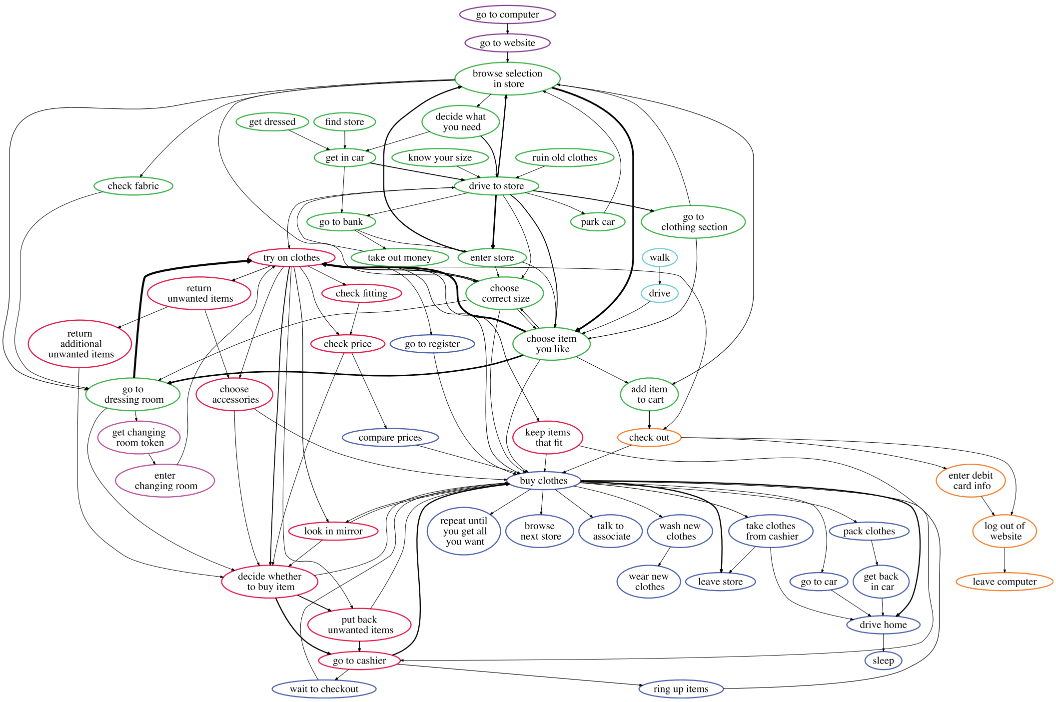

Fig. 2. Graph constructed from participants' produced activities for the event of *shopping for clothes*.

These analyses of event structure suggest a reason why studies such as Galambos and Rips (1982) failed to find evidence for the encoding of temporal structure in event knowledge. For example, one event that they used was *changing a flat tire*. Galambos and Rips chose early temporally close activity pairs such as *set the brake* and *take out spare*, and later pairs such as *remove bad tire* followed by *put on spare*. They tested the prediction that if temporal order is encoded in long-term memory for event knowledge, then both the early and late pairs would be facilitatory as compared to activity pairs that are more distant in the sequence, such as *set the brake* and *put on spare*. However, they found no such difference.

This null effect may have occurred because of large but uncontrolled differences in the strength of ordering constraints between nearby activity pairs. In some parts of the sequence, K. S. Brown et al. (in preparation) graphs show relatively weak links between activity pairs, and many potential avenues emanating from an activity node. In other cases, there is a strong edge coming out of a node, and very few nodes to which it links. In other words, an event's temporal structure can vary over its time course, and this is probably the rule rather than the exception. Some componential sequences may be constrained strongly, whereas others are constrained weakly (note that this has been explored to some extent in the AI action planning and recognition literature; Botea, Müller, & Schaeffer, 2005; Yi & Ballard, 2009). Therefore, it is possible that finding evidence for the temporal structure in event knowledge depends crucially on the specific activity pairs that are used in an experiment. Important factors appear to include the degree to which the probabilistic strength of their ordering is taken into account when the experiment is designed and analyzed, and the degree to which the experimental probe is sensitive to these probabilistic differences among stimuli.

In summary, perhaps the biggest challenge for models of event knowledge is event variability. The intuition behind frames and scripts was that there are generalizations that cut across events, and it was the goal of frames and scripts to capture these generalizations. However, the inconvenient reality is that one may go to a restaurant ten times, or a thousand times, and that script will never unfold in exactly the same way. Without disputing that there are types of events and generalizations that hold true across them, there is in reality tremendous variation in how those events are instantiated. Furthermore, the variation is not random. Some variations are correlated. For example, when you pay for your meal will depend on whether you are in a fancy restaurant versus a fast food restaurant. Whether you eat your salad before or after the main course may depend on the country in which you are dining. Some elements of events are entirely optional, whereas other elements may be quite probable. This variation also occurs at the level of temporal structure. A central conclusion from the graph theoretic analyses is that the temporal structure of events is rich and variable. One possible interpretation is that it is highly unlikely that events are represented in the human mind as a sequential linear temporal order of activities. It therefore appears to be quite challenging to construct a type of representation that could explicitly encode the temporal structure inherent in events. A slightly more radical way to state this is that explicit "event representations" do not exist. In other words, rather than using a structured representation for each type of event in long-term memory, it may be the case that the temporal structure of events is an emergent property of a computational system that implicitly represents memory for time in its processing (see Botvinick & Plaut, 2004).

3. A connectionist attractor model of event knowledge

Elman and McRae (2019) implemented a prediction-based model of event knowledge to tackle the issues of learning and processing variable, dynamic, temporal event structure. In the connectionist framework, the architecture underlying cognition involves networks of processing units (or nodes). Encoded knowledge is determined by the connection pattern among units and the learned strengths of weights between units. Thus, both memory and processing are instantiated in the weights.

The architecture of Elman and McRae's (2019) model is presented in Fig. 3. Activities are represented in terms of their components, including actions, agents, patients, recipients, instruments, and contexts (typically locations). Local representations of these concepts were used. The model was trained on sequences of activities that were not labeled explicitly as events. Elman and McRae used backpropagation through time (Williams & Zipser, 1989) to train their attractor network (i.e., the network settled through time). The

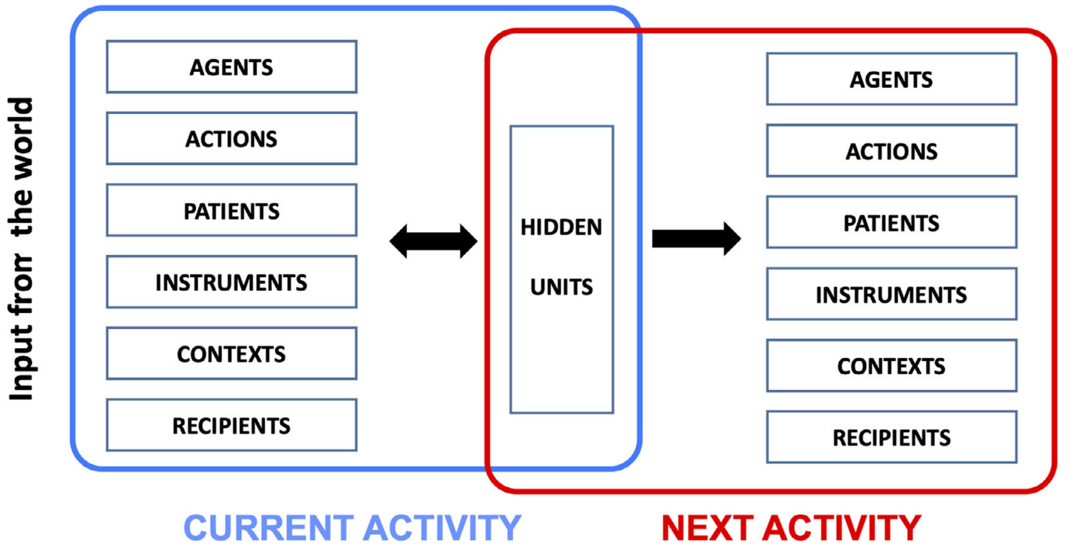

Fig. 3. Elman and McRae (2019) network architecture. Rectangles depict banks of processing units. In the Current Activity section of the network, the Agent, Action, Patient, Instrument, Context, and Recipient units are fully bidirectionally connected with the hidden units. In the Next Activity section of the network, the hidden units are fully unidirectionally connected to the right-hand-side banks of units.

left side of the network receives input from the world about the current activity. These "current activity" units are fully connected to a set of hidden (computational) units that feed back to them. These feedback connections allow the model to learn co-occurrences among the components of activities that occur in the moment. The right side of the network contains the same representational units as does the left side. The hidden units are unidirectionally connected to these "next activity" units. The purpose of this part of the network was to learn information about sequential patterns among activities by learning to predict upcoming activities.

The model processed each activity over four cycles of activation, with each cycle containing four time ticks. During cycles 1 and 2 (8 time ticks), the input was clamped on the current activity units. Activation flowed throughout the network for all time ticks. During cycles 3 and 4 (the final 8 time ticks), the input was removed so that the network was free to instantiate and perhaps complete the pattern on the current activity units, and to continue to activate the next activity units. During training, error was based on target activations for both the current and next activity units throughout cycles 3 and 4. Elman and McRae (2019) showed that when the trained model was given a partially specified current activity, it probabilistically filled in the activity's missing components based on the activities to which it had been exposed (see their fig. 2, p. 260). Thus, the network implemented pattern completion in the moment, and pattern completion through time (i.e., prediction). Furthermore, Elman and McRae showed that the model's predictions are contextually determined in that what is predicted to occur next depends not only on the current activity (as it would in a Markov chain), but also on the activities that precede it.

Also note that Elman and McRae's model is not a simple recurrent network (SRN, aka Elman network). An SRN can accomplish some aspects of what Elman and McRae's model does, but it is not an attractor network (it is a modified feed-forward network). Therefore, an SRN does not settle over time, which Elman and McRae wanted their network to do in order to illustrate temporal dynamics during the presentation of each activity. They also wanted their model to learn the co-occurrences among the components of the current activity, which an SRN is not designed to do.

One set of simulations conducted by Elman and McRae (2019) involved training a network on the participant-generated sets of activities that were described in Section 2. They used *changing a flat tire* and *going to a picnic* as examples because these events differ substantially in the degree to which their temporal structure is defined. The model learned by integrating information about the temporal order of activities and their components across the participants' productions. Elman and McRae conducted simulations in which they treated the network as though it was a participant in the activity production task. The network was seeded an initial activity, *pull car over to side of road*. Then, the most strongly predicted components were used as the next current activity, and so on. The model produced a completely sensible way of fixing a flat tire (see fig. 15, p. 275, Elman & McRae). It produced a description of the event that captures both the sequential structure of the event and also the appropriate elements of each activity. However, it was not a description that was provided by any of the participants. Instead, the network discovered what might be considered as the core consensus from the descriptions on which it was trained.

The *going to a picnic* simulation produced somewhat different results (see fig. 17, p. 277, Elman & McRae). Going to a picnic includes a number of activities that occur with reasonable frequency, although many are not necessary for having a picnic, and there is a great deal of optionality in their order (one can eat, play volleyball, and swim in any order). Because of this variability, the model produced a bare-bones picnic, consisting basically of getting there, putting out the food, eating the food, sitting, packing up, and leaving.

These simulations invite a number of conclusions regarding the form of event knowledge and how it is learned. The flexibility and variability in events argue against the idea that the human mind contains pre-specified event templates, or some type of data structure that represents events explicitly. Elman and McRae's (2019) network demonstrates the value of understanding event knowledge not by pre-specifying what an event is, but by focusing on how event knowledge might be learned, especially when people are exposed to variability. One challenge has long been how to know when and under what circumstances variability should be accommodated, and to what degree.

Finally, prediction turns out to be a powerful mechanism for learning about temporal structure. This has been known for a while now (Elman, 1990, 1995). However, some new things were learned from training the connectionist model on human data. Although any given instance of an event is necessarily linear, from the event norming data, and the many ways that even the same event can be carried out, we see that events have a much richer temporal structure than typically has been imagined. The model demonstrates that the temporal structure of events in the human mind may emerge as a consequence of prediction-based learning.

4. Conclusions

In this article, we discussed how the temporal structure of real-world events might be processed in human memory. We argued that there is substantial variability in temporal structure that occurs across instances of the same general type of event. There also is substantial variability across types of events in that some may be quite constrained whereas others may have few constraints on temporal structure. Event-predictive models of cognition such as those that are based on simple recurrent or attractor networks that feature prediction-based learning are able to deal with these sources of variability naturally so that event knowledge emerges.

Prediction is a key focus in this special issue and also has played central role in many recent models of event knowledge (Frank, Koppen, Noordman, & Vonk, 2003; Mayberry, Crocker, & Knoeferle, 2009; Modi, 2016; Reynolds et al., 2007; Takac & Knott, 2016a, 2016b; Venhuizen et al., 2019). For example, Reynolds et al. (2007) presented a model of the perception of event boundaries and the updating of event representations based on a simple recurrent network (Elman, 1990). Their work used as a starting point research on how people segment perceptual events (see Zacks & Swallow, 2007, for a review). Reynolds et al. (2007) showed that a simple recurrent network that was augmented with a gating mechanism could use prediction error signals to discover event boundaries, and they demonstrated that prediction error and event boundaries play a role in learning and updating internal representations of event knowledge. Reynolds et al. concluded that people's experience with repeated patterns in the world allows them to accurately predict upcoming stimuli within an event. Importantly, people are able to use transient increases in prediction error to identify boundaries between events (event segmentation), and this ability results in improved prediction about downstream activities.

Prediction (or expectancy generation) also has played a major role in theories and empirical investigations of language comprehension for a number of years. Furthermore, there is a great deal of evidence that people's event knowledge is a primary source of information for constructing on-line predictions of upcoming linguistic input (for reviews, see Altmann & Mirkovic, 2009; Kuperberg & Jaeger, 2016). To simulate event-based predictive language comprehension, Venhuizen et al. (2019) implemented a simple recurrent network in which they focused on the interaction between linguistic and real-world event knowledge. They used a simple recurrent network as the basis of a model that incrementally constructs rich, probabilistic situation model representations word by word. Comprehension was simulated by movement through a probabilistic situation (meaning) state space (Frank et al., 2003). Their model simulated Elman's (2009) words-as-cues approach in that each word served as a cue for traversing semantic state space in a context-sensitive manner. Venhuizen et al. showed that their prediction-based model could account for word surprisal (Hale, 2001; Levy, 2008) by constructing rich probabilistic meaning representations that support inferences that are driven by the integration of linguistic and event knowledge.

Elman and McRae (2019) also presented simulations in which the model displayed behaviors that have been characterized in human empirical work as demonstrating inferencing of unmentioned event components (Graesser, Singer, & Trabasso, 1994), the prediction of upcoming activities and their components (which may or may not ever occur or be mentioned; Metusalem et al., 2012), reconstructive memory (Bransford, Barclay, & Franks, 1972), and the ability to adapt to deviations from previously encountered sequences of activities. All of these behaviors emerged from the model's prediction-based learning mechanism. Note that although Elman and McRae view their network to be a model of event knowledge rather than of language processing, in these cases, they used it also to simulate language comprehension experiments because the influence of event knowledge often has been studied using psycholinguistic experiments. Their model did not, however, contain any mechanisms for processing, for example, passive sentences, or sentences such as "Before doing X, they did Y."

Finally, Takac and Knott (2016a, 2016b; this volume) describe models of the construction of event representations in working and long-term memory. An intriguing and important aspect of their modeling, and another way in which these models are tied to language processing, is that they tackled the problem of how elements of activities are assigned to roles such as agent and patient that play key parts in the event cognition literature, the language comprehension literature, and the event knowledge/language comprehension interface.

4.1. Events as graphs

Finally, we introduced an approach for studying event structure based on graph theory. We claim that by constructing an event network as an "ensemble" object using activity sequences from many participants, we arrive at an event description that is abstracted away from how any particular individual represents that event, while still being able to accurately capture event variability. This is a theoretical claim and a strong one. However, as only mentioned here but detailed elsewhere (K. S. Brown et al., in preparation), this is a falsifiable theory—the graphs make clear, empirically testable predictions, particularly with regard to activity centrality and the degree to which it is equally or unequally distributed across the activities in a given event. Moreover, the event graphs do reproduce a kind of dynamic flexibility. The natural way to generate activity sequences (e.g., predict the next activity) from the graph is to place a Markov process on the graph that moves from activity to activity, following directed edges, with probability proportional to edge strength. This can be done from any starting activity, so it is straightforward to complete an event after being "dropped into the middle" of it.

4.2. Causality

We end with some comments regarding causality. Humans care deeply about what will occur next, both in the short and long term. Being able to anticipate what will happen provides us with a guide to our own behavior and allows us to predict what others may do. Sometimes temporal structure is governed by causal structure. Other times, it is the result of cultural convention or habit. A model such as that of Elman and McRae (2019) definitely learns key statistical regularities, but it does not learn any explicit notion of causality (as in knowing that pulling the handle on a toilet causes water to rush through the bowl). Strong constraints on activity order do of course correlate to some extent with causal necessity. For example, when taking money out of an ATM, it is necessary to insert your bank card prior to typing in your personal identification number. (On the other hand, it may or may not be valid to think that inserting your bank card *causes* you to type in your personal identification number.) Furthermore, strong ordering constraints may reflect strong conventions, rather than causal necessity. Therefore, we view the model as a mechanism that learns statistical structure concerning the temporal nature of events, and that those statistics would be useful for humans to discover causality. That is, we do not suggest that Elman and McRae's model knows anything about causality per se, but the ability to learn varying strengths of statistical relationships provides important evidence for people to support an hypothesis that there may be a causal relationship between two activities. That said, if people learn about event structure using principles similar to those embodied in this model, then the temporal structure that is learned, including cases of near-invariant ordering constraints, could provide valuable clues for building theories of causality.

Acknowledgments

This research was supported by Natural Engineering & Sciences Research Council of Canada Discovery Grant 05652 to Ken McRae. Jeff Elman made huge contributions to the ideas presented in this article, as well as to our modeling and human empirical research reviewed herein. Sadly, he passed away before this article was published.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, *36*, 715–729.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609.

Anderson, R. C. (1978). Schema-directed processes in language comprehension. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 67–82). Boston, MA: Springer.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.

Barsalou, L. W., & Sewell, D. R. (1985). Contrasting the representation of scripts and categories. *Journal of Memory and Language*, *24*, 646–665.

Becker, R. B., Ferretti, T. R., & Madden-Lombardi, C. J. (2013). Grammatical aspect, lexical aspect, and event duration constrain the availability of events in narratives. *Cognition*, *129*, 212–220.

Black, J. B., & Bower, G. H. (1980). Story understanding as problem-solving. *Poetics*, *9*(1–3), 223–250.

Botea, A., Müller, M., & Schaeffer, J. (2005). Learning partial-order macros from solutions. In S. Biundo, K. Myers, & K. Rajan (Eds.), *Proceedings of the 15th International Conference on Automated Planning and Scheduling* (pp. 231–240). Monterey, CA.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*, 395–429.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193–209.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain & Language*, *106*, 132–143.

Butz, M. V., Bilkey, D., Humaidan, D., Knott, A., & Otte, S. (2019). Learning, planning, and control in a monolithic neural event inference architecture. *Neural Networks*, *117*, 135–144.

Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in language comprehension. *Cognitive Psychology*, *62*, 41–79.

Drummer, J., van der Meer, E., & Schaadt, G. (2016). Event-related potentials in response to violations of content and temporal event knowledge. *Neuropsychologia*, *80*, 47–55.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 195–223). Cambridge, MA: MIT Press.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582.

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*, 252–291.

Frank, S. L., Koppen, M., Noordman, L. G., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, *27*, 875–910.

Galambos, J. A., & Rips, L. J. (1982). Memory for routines. *Journal of Verbal Learning and Verbal Behavior*, *21*, 260–281.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–114.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Hale, J. T. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8). Stroudsburg, PA: Association for Computational Linguistics.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition, & Neuroscience*, *31*, 32–59.

Lancaster, J. S., & Barsalou, L. W. (1997). Multiple organisations of events in memory. *Memory*, *5*, 569–599.

Leicht, E. A., & Newman, M. E. J. (2008). Community structure in directed networks. *Physical Review Letters*, *100*, 118703.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Mandler, J. M. (1984). *Scripts, stories and scenes: Aspects of schema theory*. Hillsdale, NJ: Erlbaum.

Mayberry, M. R., Crocker, M. W., & Knoeferle, P. (2009). Learning to attend: A connectionist model of situated language comprehension. *Cognitive Science*, *33*, 449–496.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*, 545–567.

Minsky, M. (1974). *A framework for representing knowledge*. Cambridge, MA: MIT Press.

Modi, A. (2016). Event embeddings for semantic script modeling. In Y. Goldberg & D. Riezler (Eds.), *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)* (pp. 75–83). Berlin, Germany: Association for Computational Linguistics.

Murphy, G. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in netwoks. *Physical Review E*, *69*, 026113.

Norman, D. A., & Rumelhart, D. E. (1981). The LNR approach to human information processing. *Cognition*, *10*, 235–240.

Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience*, *26*, 1905–1917.

Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of Psycholinguistic Research*, *28*, 395–414.

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2007). Investigating dimensional organization in scripts using the pupillary response. *Psychophysiology*, *44*, 864–873.

Raisig, S., Welke, T., Hagendorf, H., & Van Der Meer, E. (2009). Insights into knowledge representation: The influence of amodal and perceptual variables on event knowledge retrieval from memory. *Cognitive Science*, *33*, 1252–1266.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*, 613–643.

Rosen, V. M., Caplan, L., Sheesley, L., Rodriguez, R., & Grafman, J. (2003). An examination of daily activities and their scripts across the adult lifespan. *Behavior Research Methods, Instruments, & Computers*, *35*, 32–48.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. C. (1980). Language and memory. *Cognitive Science*, *4*, 243–284.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Takac, M., & Knott, A. (2016a).Mechanisms for storing and accessing event representations in episodic memory, and their expression in language: A neural network model. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 532–537). Austin TX: Cognitive Science Society.

Takac, M., & Knott, A. (2016b).Working memory encoding of events and their participants: A neural network model with applications in sensorimotor processing and sentence generation. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2345–2350). Austin, TX: Cognitive Science Society.

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In L. R. Gleitman, & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society* (pp. 523–528). Austin, TX: Cognitive Science Society.

Van Der Meer, E., Beyer, R., Heinze, B., & Badel, I. (2002). Temporal order relations in language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 770.

Venhuizen, N., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, *56*, 229–255.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, *1*, 270–280.

Yi, W., & Ballard, D. (2009). Recognizing behavior in hand-eye coordination patterns. *International Journal of Humanoid Robotics*, *6*, 337–359.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*, 80–84.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning Memory, & Cognition*, *22*, 1196–1207.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.