TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 13 (2021) 256–298

ISSN:1756-8765 online
DOI: 10.1111/tops.12518

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 https://onlinelibrary.wiley.com/toc/17568765/2021/13/1

Tea With Milk? A Hierarchical Generative Framework of Sequential Event Comprehension

Gina R. Kuperberg[a,b]

[a]*Department of Psychology and Center for Cognitive Science, Tufts University*
[b]*Department of Psychiatry and the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School*

Received 5 November 2019; received in revised form 11 July 2020; accepted 11 July 2020

Abstract

To make sense of the world around us, we must be able to segment a continual stream of sensory inputs into discrete events. In this review, I propose that in order to comprehend events, we engage hierarchical generative models that "reverse engineer" the intentions of other agents as they produce sequential action in real time. By generating probabilistic predictions for upcoming events, generative models ensure that we are able to keep up with the rapid pace at which perceptual inputs unfold. By tracking our certainty about other agents' goals and the magnitude of prediction errors at multiple temporal scales, generative models enable us to detect event boundaries by inferring when a goal has changed. Moreover, by adapting flexibly to the broader dynamics of the environment and our own comprehension goals, generative models allow us to optimally allocate limited resources. Finally, I argue that we use generative models not only to comprehend events but also to produce events (carry out goal-relevant sequential action) and to continually learn about new events from our surroundings. Taken together, this hierarchical generative framework provides new insights into how the human brain processes events so effortlessly while highlighting the fundamental links between event comprehension, production, and learning.

Keywords: Bayesian; Late positivity; Monitoring; N400; Prediction; Predictive coding; Prediction error; Temporal receptive field

Correspondence should be sent to Gina R. Kuperberg, Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155. E-mail: gina.kuperberg@tufts.edu or gkuperberg@mgh.harvard.edu

1. Introduction

To act upon and make sense of the world around us—whether we are making ourselves a cup of tea, watching a movie, reading a book, or cleaning the fridge—we draw upon a rich store of knowledge, built over years of experience. How do we represent this knowledge in memory, and how do we exploit it during the comprehension and production of sequential action? In this review, I will argue that these questions can be fruitfully addressed by appealing to a common set of computational principles—probabilistic prediction, tracking the magnitude of prediction error, and tracking the certainty of our beliefs about the goals of other agents (comprehension) and ourselves (sequential action), at multiple time scales. I will bring these principles together within a *hierarchical generative framework,* showing how this framework can inform our understanding of the neurocognitive mechanisms engaged in event comprehension while highlighting its close relationship with the production of sequential action and learning.

2. Event representation

2.1. Core properties of an event

What is an "event"? The answer to this question will vary depending on the field of inquiry: A linguist will respond differently from a visual neuroscientist. Nonetheless, there seems to be some consensus about its core features and functional properties.

First, most would agree that events are composed of multiple different elements. In Linguistics, there is a long tradition of describing these building blocks in terms of a central action and the "roles" that different entities play around this action (Dowty, 1989; Fillmore, 1967; Gruber, 1965; Jackendoff, 1987; see Unal, Ji, & Papafragou, 2021, this issue for discussion). For example, the event, <Woman swallows tea>, describes an Action ("swallow"), an Agent (the "woman" who is carrying out the action), and a Theme (the "tea" that undergoes this action). These *semantic-thematic roles* are central to the structure of language,[1] and they can be easily and automatically identified, even when events are presented visually (Hafri, Papafragou, & Trueswell, 2013; Hafri, Trueswell, & Strickland, 2018).

A second key property of events is that they convey a *change in state in the world.* For example, the event, <Woman swallows tea>, describes a change in state of both the tea (less tea in the cup) and the woman (more tea in the woman). For events like this, which involve human agents, it is this change in state that *bridges action and perception* (Ballard, Hayhoe, Pook, & Rao, 1997; Hommel, Musseler, Aschersleben, & Prinz, 2001; Schmidt, 1975). Regardless of whether we are swallowing tea ourselves, or watching or reading about someone else swallowing tea, it is impossible to divorce the action of swallowing from the semantic features and functional properties of the drinker, who must have a mouth, and the tea, which must be liquid.

Third, most real-world events are finite in duration. Although some events last longer than others, almost all events will eventually come to an end, and these endings are marked by *end states*. For example, the event, <Woman swallows tea>, terminates with the end state, [Tea consumed]. In Linguistics, end states are considered key to the conceptual structure of the so-called bounded events (Comrie, 1976; Parsons, 1990; Vendler, 1957; see Ünal et al., 2021, this issue for discussion).[2] In the study of goal-relevant action, it has been proposed that the anticipation of end states is the key trigger of actions, ranging from simple motor movements ("ideomotor action," see Hommel et al., 2001; James, 1890/1981; Lotze, 1852; Prinz, 1987) to complex sequences that define longer-term goals and plans (e.g., Jones & Davis, 1965). Moreover, in addition to delineating the termination of events, end states also function to probabilistically constrain the set of events that can follow (see below).

Finally, and most importantly, events are inherently *dynamic* in nature (Altmann & Ekves, 2019; Neisser, 1976). All of the properties described above presuppose the passage of time: An agent's action can only bring about a change in the world, bridge action to perception, and come to an end because time marches on. Events therefore play a critical role in the mental representation of time by bridging the past to the present, and the present to the future.

2.2. Sequences of events

What one calls a single "event" versus a "sequence of events" also varies between fields. When we describe events using language, we play tricks with time, taking only a few words to convey activities that would unfold over much longer durations in the real world. For example, when referring to the events conveyed by the sentence, "The woman in the kitchen made herself a cup of tea and then cleaned the fridge," a linguist or psycholinguist might treat the contents of each clause as a single "event," and discuss how the two events, <woman made self tea> and <woman cleaned fridge>, are linked along different dimensions such as space (both events take place in the kitchen), time (the second event occurs after the first), reference (both events are carried out by the same woman), and other types of causal or motivational connections (e.g., one might speculate that the woman made herself a cup of tea to motivate herself to clean the fridge); see Zwaan and Radvansky (1998).

When we think about how visual events are linked during real-world action and perception—the focus of this review—we do not have the luxury of skipping across time. As we watch a woman make herself a cup of tea, this does not happen all at once; rather, this activity is composed of a sequence of shorter events that unfold over time, for example, <Fill kettle with water>, <Switch on kettle>, <Put teabag in cup>, <Pour hot water into cup>, etc. To link these events together, we draw upon our basic knowledge of how human agents act upon their environment, given the constraints of our own bodies, the surrounding space, and the affordances and functions of the objects around us (cf. Gibson, 1979; Glenberg, 1997). We use this knowledge, in combination with what has come before, to determine what will come next. The end state of even a single event functions as a *precondition* that constrains

the possibility and probability of what events can follow (for early discussion, see Knoblock, 1992; Schmidt, Sridharan, & Goodson, 1978; see also Botvinick & Plaut, 2004; Cooper & Shallice, 2000). For example, after observing the event, <Woman fills kettle>, our knowledge about a kettle's affordances increases the probability that the next event will be <Switch on kettle>. Additionally, the woman's position in space excludes the possibility that the very next event will show her opening the fridge if the fridge is outside arm's reach. Crucially, this spatial and functional knowledge is inherently time-dependent (see Neisser, 1976): A few moments later, it may be perfectly possible for the woman to open the fridge.

2.3. Event schemas

We also describe events in terms of our longer-term semantic knowledge that extrapolates and generalizes our experience over many encounters with similar events. This knowledge has been variously described as "schemas" (Anderson, 1978; Rumelhart, 1975), "scripts" (Abelson, 1981; Bower, Black, & Turner, 1979; Schank & Abelson, 1977), "frames" (Fillmore, 2006; Minsky, 1975), or "structured event complexes" (Grafman, 2002). For example, most of us have some intuition of what we mean by a "tea-making" schema. However, as discussed by McRae, Brown, and Elman (2021, this issue), as soon as we try to *formalize* the structure of these schemas in terms of fixed chains or hierarchies, or try to hard-wire them into our computational models, we are confronted with an inconvenient reality: There are countless possible individual events that can fall into a particular schema, and numerous different ways in which these events can be sequenced within a given schema. For example, a "tea-making" schema can comprise a sequence of single events like <Fill kettle with water>, <Switch on kettle>, <Put teabag in cup>, <Pour hot water into cup>, <Open fridge>, <Get milk>, <Pour milk into tea>, <Pick up cup>, and <Swallow tea>. Alternatively, it can comprise the (somewhat sad) sequence of <Put teabag in cup>, <Fill cup with cold water>, <Heat cup of water in microwave>, <Pick up cup>, and <Swallow tea>. Moreover, no individual event within a given sequence tells us exactly what the sequence is about. For example, the event <Open fridge> can be part of a "Tea-making" schema, a "Dinner party" schema, or a "Clean the fridge" schema. These observations suggest that rather than representing event schemas as fixed, crystallized memory structures, they must be encoded probabilistically such that the events and event sequences that are most likely to occur within a given schema *cluster* together in representational space (see McRae et al., 2021, this issue; Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013), but no single event or event sequence is bound to any given cluster. Moreover, clusters of schema-relevant events must be available to assemble and disassemble dynamically in any given situation so that we can mix and match according to our needs.

3. The challenges of event comprehension

To comprehend visual events as they unfold in real time, we must be able to quickly mobilize the vast body of event knowledge described above, and use it to make sense of

sequential actions produced by other agents. For example, imagine watching a woman in her kitchen going about her morning routine. We see her fill the kettle with water, boil the water, put a teabag into a cup, and pour hot water into the cup. We easily infer that she is making a cup of tea. As we continue watching, we understand that she is cleaning the fridge, talking on the phone, and working on her computer. Making sense of this information feels effortless. However, given how quickly each individual event unfolds before our eyes, and how frequently the woman's goals change, event comprehension presents an enormous computational challenge.

In this section, I first discuss the role of *probabilistic prediction* in allowing us to keep pace with the rapid speed at which the input unfolds. I next discuss the challenge of detecting boundaries between sequences of visually unfolding events, and summarize one proposal of how we meet this challenge: by tracking the magnitude of *prediction error*. I then consider a number of open questions in event comprehension, suggesting that these questions can be addressed within a hierarchical generative framework, which is introduced in the following section.

3.1. Probabilistic prediction

The major advantages of prediction during event comprehension are speed and accuracy. Perceptual information unfolds at a very fast rate. Moreover, it is sometimes ambiguous or incomplete. By predicting ahead, we can exploit our prior knowledge to fill in any gaps and remain one step ahead of the input. For example, as we watch the tea-making sequence described above, and we see that the tea is ready, with the woman's hand next to the cup, instead of waiting passively for the next event to become available, we can *predict* that the woman will next pick up the cup, giving us a head-start in processing predicted bottom-up information when it becomes available.

Most frameworks of event comprehension assume that these predictions are generated implicitly by an *event model*–a high-level representation of the prior context (the sequence of events observed thus far) that is held in an active state within working memory (e.g., Radvansky & Zacks, 2011).[3] It is also usually assumed that the predictions generated by an event model are further constrained by the presence of *schema-relevant* knowledge within working memory. Some models additionally propose that these implicit predictions are actively propagated down to lower levels of representation where they pre-activate upcoming information at these lower levels (*top-down predictive pre-activation*; see Kuperberg & Jaeger, 2016, Section 4 for discussion). I will come back to the distinction between implicit prediction and top-down predictive pre-activation later in the review. For now, I emphasize that, so long as these predictions are probabilistic and based on internal representations that mirror the statistical structure of the input itself (in this case, the event model and schema-relevant event clusters), then they should increase both the speed and accuracy of comprehension (see Kuperberg & Jaeger, 2016, Section 2 for discussion).

Supporting this idea, there is a large body of evidence from studies using event-related potentials (ERPs), a direct neural index of online comprehension, showing that, relative

to unpredictable events, predictable events are easier to process and generate less evoked neural activity from 300 to 500 ms after event onset—a smaller N400 response (Kuperberg, 2016; Kutas & Federmeier, 2011). This is true both when events are described in written text (e.g., Kuperberg, Brothers, & Wlotko, 2020; Kuperberg, Paczynski, & Ditman, 2011; Metusalem et al., 2012; Paczynski & Kuperberg, 2012; Van Berkum, Hagoort, & Brown, 1999), and when they are depicted visually using sequences of static images (e.g., Coderre, O'Donnell, O'Rourke, & Cohn, 2020; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002) or movie clips (e.g., Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008).

3.2. The challenge of detecting environmental change: Event boundaries

Because probabilistic prediction is only optimal if it reflects the statistical structure of the input itself, in order to continue to predict effectively, it is critical that we are able to detect systematic *changes* in the statistical structure of our environment. It would clearly be counterproductive to continue predicting tea-relevant events if the woman in the kitchen has moved on to cleaning the fridge. By quickly detecting evidence for such changes, we can disengage from our original {Woman makes herself a cup of tea} event model and start to build a new {Woman cleans the fridge} event model.

Evidence that we are able to detect this type of systematic change comes from the study of event *boundaries*. When viewing sequences of visual events, we are remarkably consistent in being able to identify the natural boundaries between them, and this is true at multiple time scales (Hanson & Hirst, 1989; Newtson, 1973; Newtson & Engquist, 1976; Zacks, Tversky, & Iyer, 2001). For example, when watching a series of events depicting a woman first making tea and then cleaning the fridge, different observers will consistently judge when tea making stops and fridge cleaning begins. Moreover, we are able to detect these boundaries as events unfold in real time (e.g., Cohn, Jackendoff, Holcomb, & Kuperberg, 2014; Hard, Recchia, & Tversky, 2011; Kosie & Baldwin, 2019; Sitnikova et al., 2008).

Detecting this type of systematic change in the statistical structure of our environment —that is, detecting the boundaries between sequences of visual events—is not trivial. This is because, as noted in Section 2.3, there is significant *variability* in what specific events or subsequences of events can belong to any given event model and its associated schema. For example, the event, <Woman opens fridge>, can be part of a {Woman makes herself a cup of tea} event model, a {Woman makes herself a sandwich} event model, or a {Woman cleans the fridge} event model.

If we were able to observe all the events together as a batch, detecting the boundaries between event models would be less challenging. This is because an event model is, in part, defined by the full set of relationships between its component events. However, during real-time comprehension, events become available sequentially over time. Therefore, in order to continue predicting efficiently, we need to be able to infer event boundaries based on limited information (see Qian, Jaeger, & Aslin, 2012 for a more general discussion of this challenge).

3.3. Prediction error as an indicator of environmental change

One proposal of how we are able to detect boundaries between sequences of events during real-time comprehension comes from Event Segmentation Theory (Kurby & Zacks, 2008; Radvansky & Zacks, 2011; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). According to this theory, we track the magnitude of the *implicit prediction error* produced by each incoming event—the degree to which that event is inconsistent with the state of the current event model. If we encounter an event that is very dissimilar to the event that we implicitly predicted, the resulting "prediction error" leads us to update the contents of working memory by disengaging from the old event model, and its associated schema-relevant event clusters, and switching to a new event model.

This theory is supported by evidence that event boundaries do indeed coincide with regions of unpredictability in the event stream: Events that occur immediately following an event boundary are generally rated as more difficult to predict than events that occur in the middle of two boundaries, and this is true at multiple time scales (Newtson, 1976; Zacks, Kurby, Eisenberg, & Haroutunian, 2011). The theory also makes intuitive sense. If, as we watch the woman make herself a cup of tea, we suddenly see her grab a sponge, this event would produce a large prediction error, providing us with clear evidence that both our current event model and associated tea-relevant clusters are no longer appropriate.

3.4. Open questions

The account outlined above, however, leaves open many questions. First, how do we initially build event models during event comprehension? For example, when we first start watching the woman in her kitchen, how do we know what schema-relevant clusters to retrieve from long-term memory to build a new {Make self a cup of tea} event model?

Second, if we encounter an event that produces a large prediction error, how do we know whether this error is large enough to disengage from our current event model and start to build a new one? Event segmentation theory proposes that event boundaries are inferred when the prediction error is large. But how large is "large"? Put another way, how is it that we are sometimes able to *refrain* from disengaging from our current event model, even when the input seems to violate prior predictions? For example, suppose that, after watching most of the tea-making sequence, we strongly predict that the woman is just about to pick up her cup and sip her tea, but instead we see her open the fridge; this unpredicted event will produce a large prediction error, but should we necessarily infer an event boundary and start to build a new event model? It may be that the woman is, in fact, opening the fridge to get milk for her tea. Moreover, if a large prediction error does lead us to detect an event boundary, why and how does this drive us to disengage from our current event model, retrieve new schema-relevant information from long-term memory, and switch to a new event model?

Third, is the detection of a large prediction error the only indicator of an event boundary? Baldwin and Kosie (2021, this issue) discuss evidence that, during visual event comprehension, we attend closely to incoming events at points in the event stream when we are most *uncertain* about what we will see next. This raises the possibility that, when we know that a particular sequence is coming to an end (e.g., after watching the entire tea-making sequence and finally seeing the woman drink her tea), we are able to anticipate the upcoming boundary. If this is the case, then can we exploit this uncertainty in the event stream to disengage from the current event model and its associated schema-relevant event clusters *before* we actually encounter the next unpredicted event?

I will argue that we can begin to address all these questions within a probabilistic *hierarchical generative framework* in which we use internal *hierarchical generative models*, together with an algorithm known as dynamic *hierarchical predictive coding,* to comprehend sequential events. This type of probabilistic generative framework has been used successfully to model many aspects of perception and cognition (Griffiths, Kemp, & Tenenbaum, 2008; see Perfors, Tenenbaum, Griffiths, & Xu, 2011 for an excellent introduction), and hierarchical predictive coding has been proposed as a way of instantiating probabilistic inference in the brain (Clark, 2013; Friston, 2005; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1997, 1999; Spratling, 2016b).

In the next section, I first introduce the broad principles of *hierarchical generative models,* and sketch out the structure of a three-level hierarchical generative model that we might engage to comprehend streams of visually unfolding real-world events. I next describe the principles of *dynamic hierarchical predictive coding*, linking these principles to psychological theories of event comprehension. In Section 5, I will return to the open questions outlined above, showing how this framework can begin to address them.

4. A hierarchical generative framework

4.1. Generative models: Structure and principles

At the heart of a generative framework is the *generative model*. This is an internal mental model that represents a subset of our knowledge that we believe is relevant to our current situation. It describes our probabilistic assumptions about how observations from the environment are caused (or "generated") by underlying hidden (latent) causes. In its simplest form, a generative model can be conceptualized as representing information at two levels: a lower level at which each unit represents a single observation, and a higher level that encodes the full set of possible causes of these observations. The parameters of this model describe the probabilistic dependencies between the information represented at these two levels. Therefore, the information at the higher level is encoded in a more abstract form that captures the higher-order constraints of information encoded at the lower level.

We can use this generative model to *infer* the most probable set of causes that generated any given observation or set of observations from the environment.[4] Each of these

causes is referred to as a *hypothesis*, and each hypothesis is held with a particular strength—degree of *belief*. Together, the full set of beliefs can be described as a probability distribution and so the sum of all beliefs must add up to 1. The statistically optimal way of inferring causes from observations is to invert the generative model by applying Bayes' Rule. For a given set of observations, Bayes' Rule can be used to shift our initial beliefs (the *prior* probability distribution) to a new set of beliefs (the *posterior* probability distribution)—a process known as *belief updating*.

As we will see, this type of two-level generative model is often insufficient to explain the complex and multidimensional structure of our environmental observations. However, multiple generative models can be linked together in a hierarchical fashion, with representations at higher levels of the hierarchy encoding information at successively higher levels of abstraction. In this type of *hierarchical generative model*, belief updating proceeds at multiple levels of the hierarchy, with the causes inferred at one level serving as the observations for the level above. Through multiple cycles of belief updating, the model should, in principle, settle on the combination of hypotheses that best *explains* the statistical structure of the input (Pearl, 1982).

Generative models of this kind are usually specified at Marr's first (computational) level of analysis (Marr, 1971). When instantiated at Marr's algorithmic or implementational levels, full "rational" Bayesian inference (cf. Anderson, 1990) is often neither tractable nor desirable. To my mind, a major advantage of this first level of description is that it encourages us to articulate our assumptions about probabilistic representations and processes in a way that can directly inform psychological theory (see Tauber, Navarro, Perfors, & Steyvers, 2017 for discussion). It is in this spirit that I now describe, in qualitative terms, the structure of a three-level hierarchical generative model that might represent our probabilistic assumptions about how other agents produce sequences of real-world visual events, and how we could invert this model to make sense of these inputs.

4.1.1. Inferring an event cluster from sequential events

To link visual events that unfold sequentially over time, a generative model would need to encode our basic knowledge about how other agents interact with their surroundings to produce these events. As discussed in Section 2.2, this knowledge is inherently grounded in the physical world. It includes our knowledge about space in relation to our bodies, as well the functions and affordances of objects in our surroundings (Gibson, 1979; Glenberg, 1997). As also discussed, this knowledge interacts continuously with the passage of time. It therefore follows that any generative model that encodes this knowledge must be *dynamic*—that is, it must represent a state space (albeit one that is complex and nonlinear) that continuously changes. The lower (first) level of the model would represent individual events as they become available, and the higher (second) level would represent the cluster of linked events that, at any given time, is most likely to have generated the full sequence of individual events that we have observed.

In this type of dynamic probabilistic generative model, inference necessarily entails an iterative process of belief updating and probabilistic prediction. When a new event is

observed at the first level, this provides input to the second level, and the event cluster that is being inferred is updated. The newly inferred event cluster, in turn, probabilistically predicts upcoming events that are possible and probable at that moment in time. As discussed in Section 2.2, these predictions will depend on the end state of the most recently integrated event, as well as on the full history of events observed before it (see Neisser, 1976; Schmidt et al., 1978 for early discussion). Note that this type of prediction is *implicit*: At any given time, the event cluster inferred inherently constrains the probability space of what can come next. This is entailed by the dynamic nature of the model.

When the next event is observed, the event cluster is updated again. The change that is induced by each incoming event can be conceptualized as the change in the probability distribution induced by this new event (the Kullback–Leibler divergence), and it can be thought of as an *implicit prediction error*. The newly inferred event cluster, in turn, implicitly predicts its next state, and another new event is observed. Thus, through iterative cycles of implicit probabilistic prediction and updating, this dynamic probabilistic generative model should incrementally infer the event cluster that is most likely to have generated the full sequence of individual events observed.

4.1.2. Inferring the goals of other agents from event clusters

If we engaged the dynamic generative model described above as we watch the woman making tea in her kitchen, we would be able to infer an event cluster that explains the sequence of the events observed, based on our spatial and functional knowledge of how we interact with our environment. However, this would not be sufficient to truly *comprehend* what was going on—this cluster would look nothing like an event model. This is because, as yet, our generative model has no access to any of our long-term schema-relevant knowledge: There is nothing to link the event cluster that is being inferred to any of our prior knowledge and experiences of making tea. Just as important, there is nothing yet to tell the model that there are alternatives to making tea! Therefore, if we continue to watch the woman, and we see her grab a sponge, despite this new event being unpredicted and producing a large implicit prediction error (because, based on our knowledge of sponges, it would be perceived as relatively dissimilar to the previous set of events), we would continue to infer one giant event cluster. There is no way for us to *use* this prediction error in combination with our prior knowledge about cleaning, to infer that we should start building a new event cluster.

The fundamental assumption that the model is currently missing is that agents produce sequences of actions to satisfy their longer-term *goal* and that these goals can change: The woman in the kitchen does not simply switch on the kettle because there is water inside it, and because it affords the function of boiling this water; she presumably carries out this action because she has the goal of making herself a cup of tea. It is this goal that drives her to generate the entire sequence of events that we observe. Therefore, to truly comprehend this action sequence (infer its underlying latent cause and explain the input), we must be able to infer the agent's goal (for early discussion, see Schmidt et al., 1978; see also Baker, Saxe, & Tenenbaum, 2009; Dennett, 1987).

To accommodate this basic assumption, we need to add a third level to our generative model where we represent our beliefs about the range of possible and probable goals that

the woman might consider as she goes about her activities in the kitchen. Each of these goals would be represented in an abstract form that would need to capture two further assumptions about how goals drive agents to produce sequences of events.

The first assumption is that, to carry out the sequence of actions required to achieve any given goal, other agents draw upon stored schema-based knowledge that is relevant to that goal. For example, to achieve the goal, {Make self a cup of tea}, we assume that the woman in the kitchen must draw upon her stored tea-relevant knowledge. Obviously, as comprehenders, we do not have access to exactly the same body of knowledge as this woman. However, we can associate each of her possible goals with a subset of our own probabilistic knowledge that we believe is relevant to her. Given that we do not know anything about this particular woman, this would reflect our assumptions about an "average" woman's schema-relevant knowledge. The parameters of our generative model would specify how each goal, represented at the third level of the generative hierarchy, causes/generates a set of schema-relevant event clusters represented at the second level, each with different likelihoods. For example, based on our assumptions about the average tea drinker, the goal, {Woman makes self a cup of tea}, would generate clusters that encode highly likely tea-relevant events (e.g., <Brew tea in hot water>), as well as less likely tea-relevant events (e.g., <Pour milk into tea>). However, this goal would be very unlikely to generate clusters that encode events like <Grab a sponge>.

The second assumption is that the sequence of individual events generated by any given goal will come to an end; that is, we know that a goal's representation is inherently finite with an ending that is marked by an *end state*. For example, a possible end state of the goal {Make self a cup of tea} might be [Tea consumed] (highly likely) or [Tea steeped] (less likely), but it is unlikely to be [Water boiled]. Goal end states are usually conceptualized as the desired future state of affairs that is associated with a goal's fulfillment (e.g., Jones & Davis, 1965). They are thought to be an inherent part of how goals are represented, and, just like the end states of individual events (see Section 2.2), they set preconditions over future goals (see discussion by Cooper, 2021, this issue). Our generative model would therefore need to specify, once again probabilistically, the range of possible end states associated with each goal.

Together, these three levels are linked to form a hierarchical generative model. Inference proceeds at both the highest (third) and middle (second) levels of the model as each incoming event becomes sequentially available to the lowest (first) level. At any given time, the second level of the model infers the event cluster that is most likely to have generated the sequence of individual events observed thus far—the *current event model*—while the third level of the model infers the goal that is most likely to be generating the event model inferred at the second level (see Fig. 1).

4.2. Hierarchical dynamic predictive coding

The description above focused on the basic structure of a three-level hierarchical generative model that describes our probabilistic assumptions about how visually observed real-world events are produced by other agents. To understand how our brains might

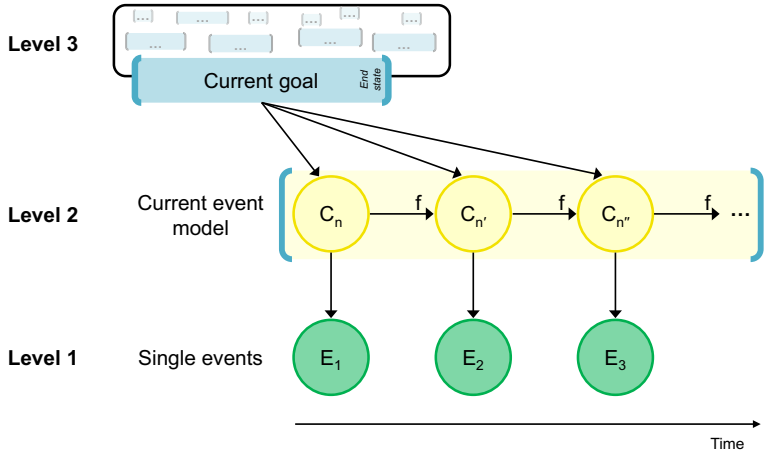

Fig. 1. The structure of a three-level hierarchical generative model that represents a comprehender's probabilistic assumptions about how other agents produce sequences of real-world visual events. The highest level of the hierarchy (level 3) represents the comprehender's probabilistic beliefs about possible goals that could be generating the observed set of events. Each goal is probabilistically associated with an end state, as well as a range of possible and probable schema-relevant event clusters, which are represented at the middle level of the hierarchy (level 2). The event model is a dynamically changing cluster, C, of linked events that represent the comprehender's understanding of the full sequence of events observed thus far, and what she believes will happen in the future. At any given time, this event model probabilistically generates the next event in the sequence, E_n based on a function, $f(E_{1:n-1})$, which is determined by the comprehender's spatial and functional knowledge of how agents interact with their environment (note that these probabilistic predictions will also be influenced by schema-relevant information). These predicted upcoming events are represented at the lowest level of the hierarchy (level 1). For further explanation, see Section 4.1 of the manuscript.

actually *use* this hierarchical generative model as an inference engine during real-time event comprehension, we need an algorithm that can carry out (or approximate) iterative belief updating by passing information from one level of the cortical hierarchy to another in both a top-down and bottom-up fashion.

While there are several different algorithms that can, in principle, carry out or approximate Bayesian inference in the brain (Aitchison & Lengyel, 2017), the best known of these is *hierarchical predictive coding*. This algorithm was first instantiated in a neural network to simulate extra-classical receptive-field effects in the visual system (Rao & Ballard, 1999; see also Lee & Mumford, 2003; Mumford, 1992; Spratling, 2008). It was later expanded into a more general theory of how hierarchical probabilistic inference is carried out across multiple domains of perception and cognition, and how information is passed up and down the cortical hierarchy (Clark, 2013; Friston, 2005; see also Spratling, 2016b).

Predictive coding is not only generative in a theoretical sense; it is *actively* generative (cf. Hinton, 2007; Hinton, Dayan, Frey, & Neal, 1995) — that is, within each two-level generative model, information represented at the higher level (level 2) is actively

propagated down, through feedback connections, to reconstruct activity at the lower level (level 1) (Rao & Ballard, 1999). In dynamic predictive coding, as originally instantiated by Rao and Ballard (1997) through the application of an extended Kalman filter (see also Friston, 2005; Rao, 1999), the state at the higher level is assumed to change continuously over time. It therefore continually generates implicit probabilistic predictions about its future state, and it is these *implicit predictions* that are propagated down as reconconstructions to the lower level of the hierarchy, thereby changing the state of activity at the lower level *before* the next input from the environment becomes available (*top-down predictive pre-activation*).

The reconstructed activity at the lower level (level 1) is subtracted from the state that is induced when new environmental input actually appears, and only the difference in activity (*observed–predicted*) is passed back up to the higher level (level 2) via feedforward connections. I will refer to the difference between the new bottom-up input and the top-down reconstruction of input at level 1 as the *first-level bottom-up prediction error*.[5] When this bottom-up prediction error reaches the higher level (level 2), it induces a change in its state. The shift from the old to the new state is the *implicit prediction error*. As each new input becomes available from the environment, this process is repeated until the magnitude of the bottom-up prediction error is minimized. In this way, the algorithm either approximates (Rao & Ballard, 1999) or, under certain assumptions, carries out Bayesian inference (Friston, 2005; Spratling, 2016a).

In hierarchical predictive coding, exactly the same process proceeds at all levels of the generative hierarchy. For example, representations at a still higher level (level 3) of the hierarchical generative model are propagated down to level 2 where they generate their own probabilistic reconstructions. The degree to which the state at level 2 is updated on any given cycle of belief updating will therefore reflect a compromise between any pre-activation it receives from level 3 and the bottom-up prediction error it receives from level 1. Finally, the top-down reconstruction at level 2 is subtracted from the newly inferred state at level 2. If this yields a difference, then the resulting *second-level bottom-up prediction error* is, in turn, passed up to level 3 of the hierarchy to update beliefs over the hypotheses that are generating the full set of inputs observed. Thus, because of how each level is linked to the level above and the level below, information flows up and down the hierarchy until error is minimized across the entire generative model. By minimizing error, the model should settle on a state—an *interpretation*—that best explains the full set of observed inputs.

4.3. Mapping computational principles on to cognitive representations and processes

The description of hierarchical predictive coding offered above is somewhat abstract. I now consider how this algorithm would carry out inference across the three-level hierarchical generative model sketched out in Section 4, linking this to the representations, constructs, and cognitive mechanisms that are typically discussed in psychological models of event comprehension.

As discussed in Section 3.1, most psychological models of event comprehension appeal to the idea that an ongoing interpretation—an *event model*—is represented in an active state within working memory. Within this hierarchical generative framework, this active state of working memory corresponds to the current state of activity at the *second (middle) level* of the three-level hierarchical generative model described in Section 4. The *event model* itself is the dynamically evolving event cluster that is being inferred at any given time. It reflects our understanding of what we have observed, what we are currently observing, and what we believe that we are about to observe in a given event sequence. For example, as we watch the woman in her kitchen, at a particular moment in time, our event model might constitute some structured representation of what we understand about "this woman in her kitchen making herself a cup of tea and currently opening the fridge," and it would also encode *implicit probabilistic predictions* about its future state (the woman reaching for a range of possible items that can be kept in a fridge, as well as other possible upcoming events). These implicit predictions reflect our beliefs about the possible and probable events that we will next encounter, given our spatial knowledge and the affordances of surrounding objects. Note that they correspond to the type of implicit predictions that are instantiated by *recurrent neural networks*, which are commonly used to model both event comprehension (e.g., Butz, Bilkey, Humaidan, Knott, & Otte, 2019; Elman & McRae, 2019; Franklin, Norman, Ranganath, Zacks, & Gershman, 2020; Hanson & Hanson, 1996; Rabovsky, Hansen, & McClelland, 2018; Reynolds, Zacks, & Braver, 2007) and event production (e.g., Botvinick & Plaut, 2004; Cooper, Ruh, & Mareschal, 2014).[6]

In addition, at any given time, the third level of the hierarchical generative model, which represents our beliefs about the goals of other agents, is actively generating probabilistic top-down predictions of event clusters at the second level of the hierarchy within working memory. Recall that the hierarchical generative model is structured such that each possible goal probabilistically generates a range of event clusters, each with different likelihoods, and that, together, these event clusters correspond to the long-term schema-relevant knowledge that is associated with that goal. The assumption here is that this schema-relevant information is latent within long-term memory, but linked to goal representations so that it can be proactively retrieved (activated within working memory) as needed during comprehension. As a result, at any given time, the implicit predictions generated by the event model represent the intersection between (a) the set of possible and probable upcoming events, given our spatial knowledge and the affordances of surrounding objects, and (b) the set of schema-relevant event clusters that have been activated/retrieved, based on our beliefs about the current goal.

In dynamic hierarchical predictive coding, these implicit predictions are actively propagated down to the first level of the hierarchy, thereby probabilistically pre-activating representations of upcoming single events—*top-down predictive pre-activation* (see Kuperberg & Jaeger, 2016, Section 4). The degree of pre-activation will depend on the predictive constraint of the current event model. For example, a weakly constraining event model would generate widely dispersed weak pre-activation over multiple possible upcoming event representations, whereas a highly constraining event model would

generate strong and focused activity over one highly likely upcoming event representation. The main advantage of this type of top-down predictive pre-activation, over the type of *implicit prediction* described above, is that it can give us even more of a head-start in processing new input when it becomes available. So long as the structure of the generative model also reflects our own comprehension goals (see Section 5.6 for discussion), and so long as probabilistic pre-activation is based on the correct event model, then new inputs to the first level of the hierarchy, should, on average, be supported by this prior pre-activation, and processing should proceed more accurately and efficiently than if we did not pre-activate at all.

When a new event is observed at the first level of the hierarchy, the difference between the activity that it induces and the activity that was predictively pre-activated constitutes the first-level bottom-up prediction error induced by this new event. Note that the use of the term "prediction error" here does *not* imply a "prediction violation" or "error" in the colloquial sense: The first-level prediction error produced by an incoming event simply reflects the new information provided by observing this event—the information that was not reconstructed before it was observed (see also footnote 5). The magnitude of the prediction error produced by the incoming event is proportional to the likelihood of its prior pre-activation (see Brothers & Kuperberg, 2020), and it reflects the amount of "work" required to initially retrieve/access this event representation. Neurophysiologically, the difficulty of retrieving this event information, and therefore the magnitude of the first-level prediction error, is thought to be reflected by the amplitude of the N400 ERP component: The more predictable the event, the smaller (less negative) the N400 (see Kuperberg, 2016; Kuperberg et al., 2020 for discussion), although note that the N400 produced in response to visual events has a more anterior scalp distribution and a more extended time course than the N400 produced by single words during language comprehension (e.g., Coderre et al., 2020; Cohn et al., 2012; Sitnikova et al., 2008; West & Holcomb, 2002).[7]

When this unpredicted event information (the informational content of the bottom-up prediction error) reaches the second level of the hierarchy, it will update the current event model, inducing a shift in its state. The magnitude of this shift is the *implicit prediction error* and corresponds to the "prediction error" that is referred to in some models of event comprehension (e.g., Radvansky & Zacks, 2011; Reynolds et al., 2007; see also Rabovsky et al., 2018). This may or may not be equal in magnitude to the first-level bottom-up prediction error described above (see Kuperberg et al., 2020 for discussion).

To sum up, at any given time, working memory, represented at the second level of the generative hierarchy, includes the current event model as well as its implicit probabilistic predictions about its future state. These implicit predictions are determined by both our spatial and functional knowledge about how we interact with our surroundings, as well as the schema-relevant event clusters that have been proactively retrieved, based on our beliefs about the current goal. These implicit predictions, in turn, lead to the probabilistic top-down pre-activation of the next event in the sequence at the first level of the hierarchy, see Fig. 2.

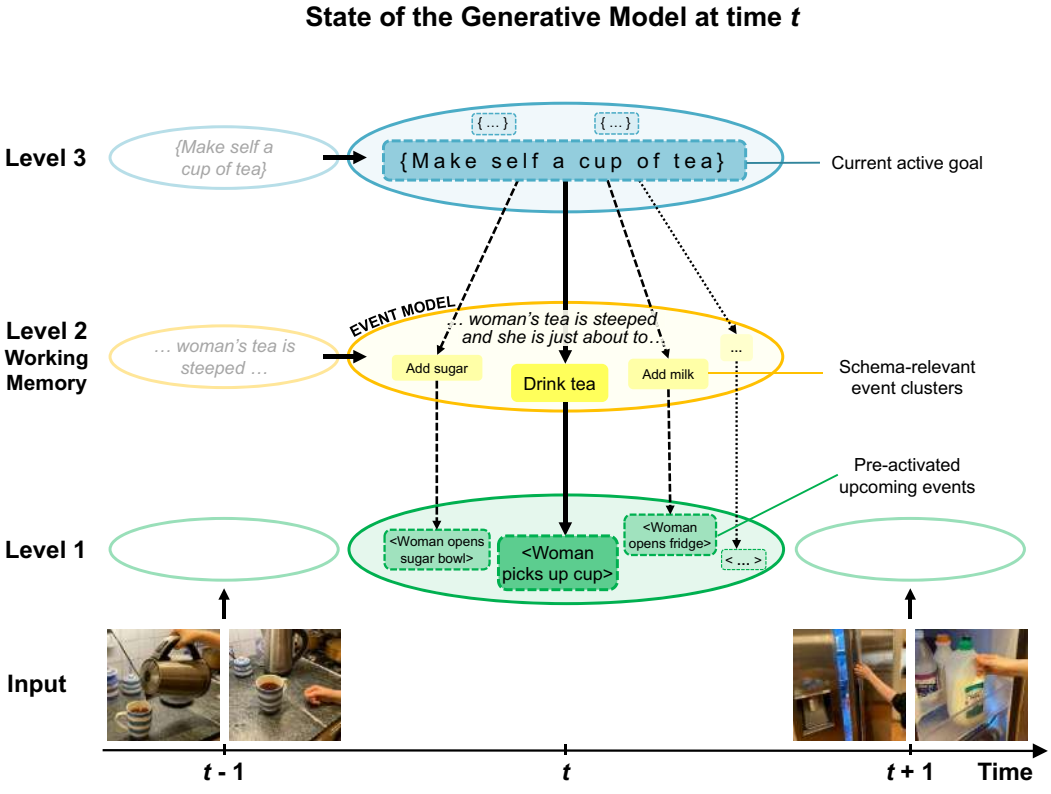

Fig. 2. The state of an actively generative three-level hierarchical model at time t, as we watch the tea-making sequence. Activity at the middle (second) level of the generative hierarchy represents the active state of working memory where the current event model is being inferred. This event model is a dynamic representation that reflects our understanding of the sequence of events that we have observed (a woman making tea in her kitchen, who has just poured hot water into a cup, resulting in a cup of steeped tea), as well as our probabilistic beliefs about what we are about to observe. At time *t*, these probabilistic beliefs reflect the intersection between (a) the set of possible and probable upcoming actions, given that the woman's hand is positioned near the cup and the tea is freshly steeped, and (b) the set of schema-relevant event clusters that have been retrieved from long-term memory, based on our higher-level beliefs about the woman's overall goal. For example, a strong belief that the woman's overall goal is {Make self a cup of tea} may lead to the top-down activation of schema-relevant clusters that encode sequences such as "Drinking tea" (highly likely), "Adding milk," or "Adding sugar" (less likely). These implict probabilistic predictions within working memory result in the top-down probabilistic pre-activation of the next event in the sequence at the lowest level of the hierarchy (level 1), for example, <Woman picks up cup> (highly likely), <Woman opens fridge>, or <Woman opens sugar bowl> (less likely). At time $t + 1$, the next event is observed, <Woman opens the fridge>. The resulting bottom-up prediction error would be passed up to the middle level of the hierarchy (working memory), leading to an update of the event model (not depicted). For further explanation, see Section 4.3 of the manuscript.

Having now described the computational principles of generative models and hierarchical predictive coding, we are now in a position to revisit the open questions outlined in Section 3.4 and see how these principles can inform our understanding of the neurocognitive mechanisms engaged in event comprehension. I turn to this in the next section.

5. Engaging hierarchical generative models to comprehend sequences of events

5.1. Homing in on an event model

The first question we asked was, when we first start to observe a sequence of events, how do we build a new event model, and how do we know what schema-relevant clusters to retrieve from long-term memory? To see how this framework can address this question, let us return to the point when we first start watching the woman in her kitchen.

At this stage, we will be uncertain about the woman's overall goal. Within a probabilistic framework, our uncertainty about the underlying latent cause of our observations is known as *expected uncertainty* (or *estimation uncertainty*; Dayan & Yu, 2003; Yu & Dayan, 2005). A basic principle of Bayesian inference is that the greater our prior expected uncertainty, the more we update our beliefs upon encountering new unpredicted input (*expected surprise*). Therefore, towards the beginning of the event sequence, the rate at which we update our beliefs will be high, enabling us to home in rapidly on the goal that is generating the event sequence we observe.

Within a predictive coding framework, our high initial uncertainty about the woman's goal means that top-down pre-activation to the lower levels of the generative hierarchy will be minimal. Therefore, when we encounter the first event of the sequence (<Woman fills kettle with water>), it will produce a large first-level bottom-up prediction error (its information has not been pre-activated). This unpredicted event information is then passed up to the second level of the hierarchy (working memory), where, based on our knowledge about the woman's position in space and the functional affordances of a kettle, it implicitly probabilistically predicts its future state. In this way, as each incoming event becomes available in real time, through cycles of probabilistic prediction and belief updating (as described in Section 4.1.1), we start to build a new event model.

At any given time, this event model, in turn, provides new information that is passed up to the third level of the hierarchy, which represents our beliefs about the woman's goals. Although on any given cycle of belief updating at the second level, the amount of information that is passed up to the third level is relatively small, as we observe more events, we will converge with increasing certainty on this goal. This will, in turn, lead to the generation of increasingly strong top-down pre-activation over tea-relevant event clusters, corresponding to the proactive retrieval of this information from long-term memory into working memory at the second level of the hierarchy. These active schema-relevant clusters will further constrain the implicit probabilistic predictions that the event model generates about its future state at any given time. For example, when we see the woman open the fridge, the prior activation of tea-relevant event clusters within working memory will increase the probability that we will next see her get milk (rather than, say, orange juice) from the fridge.

The increasingly strong implicit predictions generated by the event model within working memory will, in turn, generate increasingly strong top-down pre-activation over upcoming individual event representations at the first level of the hierarchy. As a result,

as each new consistent incoming event is observed at the first level, it becomes progressively easier to access/retrieve it (i.e., its processing is facilitated), producing a progressively smaller first-level prediction error. Evidence for this type of progressive facilitation in processing incoming events comes from a study showing that, during sequential visual event comprehension, the amplitude of the N400 produced by the first event is large and becomes progressively smaller as the sequence unfolds (Cohn et al., 2012).

5.2. *Reactively detecting event boundaries by tracking prediction error*

The second question we asked was, how large must a prediction error be to infer the presence of an event boundary? We also asked why and how the detection of a large prediction error would lead us to disengage from our current event model (and associated schema-relevant clusters) and begin retrieving new schema-relevant clusters from long-term memory in order to build a new event model.

To illustrate the answer to these questions, imagine that we have observed almost all the events in the tea-making sequence, and we see the woman approach her freshly brewed cup of tea. At this point, we are nearly certain that her overall goal, represented at the third level of the hierarchy, remains {Make self a cup of tea}. Our current event model, represented within working memory at the second level of the hierarchy, strongly predicts its upcoming state, which, in turn, generates strong top-down pre-activation over the most likely upcoming event at the first level—<Woman picks up cup>.

First imagine that instead of seeing the woman pick up the cup, we see her open the fridge. This will induce a large bottom-up prediction error at the first level of the hierarchy, and a large implicit prediction error at the second level (a large shift as we update our current event model with this unpredicted event). Despite this, we do not infer that there has been any change in the woman's overall goal, and we do not disengage from our current event model. Now imagine that instead of seeing the woman pick up the cup as predicted, we see her grab a sponge. This time we *do* infer that there has been a change in her overall goal, and we do disengage from the current event model. At a computational level, the key difference between these scenarios is that, in the first case, the newly inferred event model falls within the range of likely event clusters that can be generated by the {Make self a cup of tea} goal. Therefore, after shifting to the unpredicted event, <Woman opens fridge>, the event model that is being inferred at the second level of the generative hierarchy can still be *explained* by this overall goal. In contrast, in the second case, after integrating <Woman grabs sponge>, the full event model cannot be explained, and this leads to a redistribution of our beliefs about the woman's underlying goal at the third level of the generative hierarchy.

This exemplifies a fundamental principle of Bayesian inference known as the Bayesian Ockham's razor (MacKay, 2003, Chapter 28): Although we try to explain incoming data as simply as possible, if we encounter new information that is very unlikely given these prior assumptions, we will always infer (retrieve or learn) the hypothesis that assigns the highest likelihood to the data (see also Shin & DuBrow, 2021, this issue for a discussion in relation to classical rational models of categorization, cf. Anderson, 1991 and Sanborn,

Griffiths, & Navarro, 2010). In the present case, this is a {Woman cleans} goal, which better explains <Woman grabs sponge> than the originally assigned {Woman makes self a cup of tea} goal. The Bayesian Ockham's razor is instantiated by a class of nonparametric Bayesian models known as *Dirichlet process infinite mixture models*. In recent work, Franklin et al. (2020) show that this type of model is indeed able to infer event boundaries in response to highly unlikely events that produce a large implicit prediction error—a large shift from an old to a new state in a recurrent neural network.

More generally, a large prediction error that leads us to infer that the statistical structure of the environment has fundamentally changed is known as *unexpected surprise* (Dayan & Yu, 2003; Yu & Dayan, 2005). Unexpected surprise is associated with a *reactive re-allocation of attention* (Yu & Dayan, 2005; see also Feldman & Friston, 2010). It can lead to a number of different consequences that depend on the structure of the agent's generative model.[8] In the situation described above, the unexpected surprise induced by the unexpected <Woman grabs sponge> event forces us to "go back in time" and retroactively re-evaluate the belief that we had about the woman's goal before we observed the unexpected event. Moreover, the large redistribution of belief over goals at the highest level of the hierarchical generative model (large *Bayesian surprise*, see Baldi & Itti, 2010) also leads to a retroactive redistribution of beliefs at lower levels of the generative hierarchy to ensure that the input is explained across the entire model (Pearl, 1987).

Within a dynamic predictive coding framework, this type of retroactive redistribution of beliefs at the highest level of the generative hierarchy is triggered by a large *second-level* bottom-up prediction error, and the retroactive redistribution of activity at lower levels of the hierarchy is driven by retroactive top-down feedback activity (see Friston, 2005; Lee & Mumford, 2003). Prior to observing the highly unexpected event, <Woman grabs sponge>, our high prior certainty about the woman's goal had led to the proactive retrieval of a full range of tea-relevant event clusters within working memory at the second level of the hierarchy. However, when <Woman grabs sponge> was incorporated into the event model, the newly inferred event model failed to match any of these active clusters. In other words, it *conflicted* with the contents of working memory, producing a second-level prediction error, which, when passed up to the third level, led to the redistribution of belief over goals. The resulting retroactive top-down feedback to the second level of the hierarchy resulted in (a) a *suppression* of the current event model and its associated tea-relevant event clusters—a *disengagement* from the current event model and its associated schema-relevant clusters within working memory, and (b) an *enhancement* of activity over cleaning-relevant clusters—the *top-down retroactive retrieval* of new schema-relevant information from long-term memory. This, in turn, set the stage for building a new event model within working memory.

Evidence for this type of late retroactive processing comes from ERP studies showing that, in addition to the N400 component, unexpected events can sometimes also evoke a late frontally distributed positive-going ERP waveform that is visible on the scalp surface between 600 and 1,000 ms. As discussed earlier, the N400 is evoked by *any* semantically unexpected input. Within this hierarchical generative framework, its amplitude reflects

the magnitude of the first-level bottom-up prediction error that is produced by incoming information about a new event that has not already been pre-activated. In contrast, a *late frontal positivity* is only produced when new unpredicted input triggers a late, high-level re-interpretation of the prior context (see Brothers, Wlotko, Warnke, & Kuperberg, 2020; Kuperberg et al., 2020, for a detailed discussion within this generative framework). During language comprehension, this can occur when, following a highly constraining context, an unexpected incoming word leads us to revise our earlier high-certainty beliefs about the current discourse model, analogous to revising our high-certainty beliefs about the woman's tea-making goal after seeing her grab a sponge (e.g., Brothers, Wlotko, et al., 2020; DeLong, Quante, & Kutas, 2014; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kuperberg et al., 2020; Van Petten & Luka, 2012). Following low constraint contexts, a similar *late frontal positivity* is also elicited by highly informative words that also trigger the retrieval of new schema-relevant clusters from long-term memory that lead to a retroactive interpretation of the prior context (e.g., Brothers, Greene, & Kuperberg, 2020; Chow, Lau, Wang, & Phillips, 2018; Freunberger & Roehm, 2016).[9]

5.3. Proactively detecting event boundaries by tracking uncertainty

The third question we posed in Section 3.4 was whether it is possible to infer an event boundary in the absence of a very large prediction error. In their contribution to this Special Issue, Baldwin and Kosie (2021, this issue) discuss evidence that, as we watch sequential images of everyday events, our gaze times ramp up *before* we encounter unpredicted events at points in the sequence when we are most *uncertain* about the upcoming input. This suggests that under some circumstances, we *predict* upcoming event boundaries (Hard, Meyer, & Baldwin, 2019; Hard et al., 2011; Kosie & Baldwin, 2019). If this is the case, then how do we detect this rise in uncertainty, and are we able to exploit it to disengage from our current event model before we observe the upcoming unpredicted event?

Again, this probabilistic framework provides a principled answer to these questions. This is because, within this framework, we continually track our *expected uncertainty* about the woman's goals, represented at the highest level of the generative model. As discussed in Section 4.1.2, each goal is represented with its own *end state*. Therefore, after watching the entire tea-making sequence and seeing the woman finally drink her well-earned cup of tea, our expected uncertainty about her next goal will rapidly rise.[10] As a result, when we see the next unpredicted event, we will be *prepared* to shift our high-level beliefs so that we can rapidly home in on the new goal that will drive her to produce the next sequence of events we observe (see Section 5.1).

Importantly, because our expected uncertainty about the woman's upcoming goal starts to rise *before* the next unexpected event is actually observed, this uncertainty-driven detection of event boundaries is inherently *proactive*. It has been proposed that our ability to track expected uncertainty over time provides a normative account of *proactive attention* (Dayan, Kakade, & Montague, 2000; Yu & Dayan, 2005; see also Pearce & Hall,

1980): The more uncertain we are about an upcoming input in the perceptual stream, the more attention we allocate to the environment just *before* we encounter this new input.

This proactive allocation of attention associated with expected uncertainty offers several advantages over the *reactive* re-allocation of attention and reactive processing associated with unexpected surprise described above. First, within a predictive coding framework, the rapid rise in uncertainty as we reach a particular goal's end state will lead to a reduction in top-down activation at the second level of the hierarchy; that is, it will lead us to *proactively disengage* from the current event model and its associated schema-relevant clusters *before* we actually encounter the next unpredicted event. This means that when we do encounter the unpredicted event at the beginning of a new sequence, we do not need to devote unnecessary resources to going back in time and retroactively suppressing our event model and schema-relevant clusters within working memory, as discussed above.

Second, the pre-allocation of attention to our environment as we reach a goal end state offers us another opportunity: We can actively look for clues in our environment to tell us what schema-relevant clusters to retrieve next—a process that Baldwin and Kosie (2021, this issue) refer to as *information optimization*. For example, as we watch the woman take her last sips of tea, we might follow her gaze toward a sponge. This would allow us to proactively narrow in on a space of possible cleaning-related goals. Computationally, this is known as *active sensing* (Friston, Adams, Perrinet, & Breakspear, 2012; Friston et al., 2015; Yang, Wolpert, & Lengyel, 2016). In a system with limited resources, like the brain, active sensing provides us with an optimal way of gathering information from the environment that is deemed most likely to be useful in the future (Chater, Crocker, & Pickering, 1998; MacKay, 1992; Nelson, 2005).

5.4. Monitoring the dynamics of the broader environment: From unexpected to expected surprise

To sum up, this probabilistic hierarchical generative framework offers two mechanisms by which we can infer boundaries as visual events unfold sequentially in real time. It also offers insights into how we are able to use this information to disengage from our current event model and build a new event model within working memory.

First, as proposed by event segmentation theory, we can track the magnitude of the implicit prediction error that is produced when each incoming event shifts the state of the current event model—the difference between the prior and the new state of the event model (see Reynolds et al., 2007). Importantly, this probabilistic framework offers a principled explanation for how large this prediction error must be to infer an event boundary and switch to a new event model: The incoming event must produce *unexpected surprise*—it must be unlikely enough to override our prior certainty of the goal (latent cause) that we believe is generating the current event model (see also Franklin et al., 2020; Gershman, Radulescu, Norman, & Niv, 2014). Within a predictive coding framework, the incoming event must *conflict* with information that is already active within working memory, thereby producing a *second-level* bottom-up prediction error, which drives a late top-

down retroactive disengagement from the current event model and retrieval of new schema-relevant event clusters.

The second mechanism by which we can infer an event boundary and switch to a new event model is to continually track our uncertainty about other agents' goals. As we reach a goal's end state, the rapid rise in expected uncertainty about the next goal will lead us to proactively disengage from the current event model and schema-relevant clusters within working memory (reduced top-down activation from the third level of the hierarchy). Moreover, when we do encounter the next unpredicted event (expected surprise), our prior uncertainty will drive us to rapidly and incrementally home in on a new event model, by incrementally retrieving new schema-relevant clusters and inferring the new goal that is driving the new sequence of events.

These two mechanisms are distinct. The first is driven by a large prediction error (unexpected surprise) and is *reactive* in nature, entailing a *re-allocation* of attention. The second is driven by uncertainty and is *proactive* in nature, allowing us to *pre-allocate* attention to environmental inputs, and prepare for the upcoming expected surprise at the event boundary. As discussed above, of these two mechanisms, the latter is clearly preferable: It is far more efficient to proactively disengage from the current event model in response to an expected rise in uncertainty than to retroactively disengage and play "catch up" following unexpected surprise). One way in which we can reduce the chance of unexpected surprise, and increase the chance of expected surprise, is to monitor the broader dynamics of our environment and adapt our "mode of processing" accordingly.

To illustrate this, let us return to the moment of unexpected surprise when we saw the woman in the kitchen grab a sponge, just when we had predicted that she would pick up her cup and sip her tea. As we now watch her clean the fridge, we start to build a new {Woman cleans the fridge} event model. However, she then briefly returns to her cup of tea (once it has cooled) to take the first sip. Even though our {Woman makes self a cup of tea} event model, and its associated tea-relevant event clusters, are no longer active in working memory, we should still be able to rapidly retrieve this information from long-term memory, enabling us to incorporate <Woman sips tea> into that event model. This is because we have a prior for retrieving *recently used* clusters, which allows us to easily switch between event models during comprehension (see Collins & Frank, 2013; see also Ericsson & Kintsch, 1995 for a discussion of the role of "working long-term memory" in comprehension).

Now imagine that as we continue to watch, we see the woman switch between tea drinking, phone answering, and other tasks. Although this would lead to repeated unexpected surprise (see Yu & Dayan, 2005), by tracking the broader speed at which the environment is changing (tracking its *volatility*; see Behrens, Woolrch, Walton, & Rushworth, 2007; Nassar, Wilson, Heasly, & Gold, 2010; Nassar et al., 2012; see O'Reilly, 2013 for discussion), we should be able to turn this unexpected surprise into expected surprise—that is, by inferring that the environment is highly volatile, our *expected uncertainty* about the woman's goal at any given time should increase. As a result, we will attend more proactively to our surroundings (we may even actively search for potential switch cues—active sensing). Within a predictive coding framework, this will, in turn, lead to reduced

top-down pre-activation of specific upcoming events (see Brothers, Dave, Hoversten, Traxler, & Swaab, 2019; Brothers, Swaab, & Traxler, 2017 for consistent evidence). This framework therefore provides a principled explanation for how we are able to transition from a more reactive to a more proactive mode of processing (see Braver, 2012), efficiently allocating resources so that our mechanism of comprehension is calibrated to the broader dynamics of the input.

5.5. Expanding the generative hierarchy

To illustrate how the principles of hierarchical generative models and predictive coding can inform our understanding of the neurocognitive mechanisms of event comprehension, I have appealed to a relatively simple three-level hierarchy in which goals directly generate event clusters, which, in turn, generate single events that unfold sequentially over time. However, during event comprehension, we will sometimes need to represent the goals of other agents that stretch over shorter or longer time spans. For example, as we see the woman in the kitchen opening the fridge, we may be able to infer that her shorter-term *subgoal* is {Add milk to tea}. And as we continue to watch her go about her routine, we may infer that her longer-term goal is {Make breakfast}.

In the literature on action planning, goals at successively longer time scales are often represented at successively higher levels of hierarchical structure (e.g., Barker & Wright, 1954; Bower et al., 1979; Cooper & Shallice, 2000; Knoblock, 1992; Miller, Galanter, & Pribram, 1960; Newell & Simon, 1972; Norman & Shallice, 1986; Schmidt, 1976). Fig. 3 gives an example of this type of *hierarchical action plan* associated with the goal, {Make self a cup of tea}. In this hierarchical structure, representations at lower levels are *embedded* within representations at higher levels and it is only possible to transition to higher levels of the hierarchy when the end states at lower levels have been satisfied (see Zacks & Tversky, 2001 for discussion).[11] Note also how the structure of "goals" and "subgoals" at higher levels echoes the structure of individual "events" represented at the first level: Even though each represents information at a different time scale, each has its own end state (indeed, as noted at the outset of this review, a psycholinguist may well describe the goal, {Woman makes self a cup of tea}, as a single "event").

Fig. 3 shows the woman's tea-making action plan "after the fact." Of course, when we first start watching her activities, we have no way of knowing what this plan will be. However, by *further expanding* the probabilistic hierarchical generative model sketched out above, and using this expanded model to track both the magnitude of prediction error and our expected uncertainty over the woman's goals and subgoals *at multiple time scales*, we should, in principle, be able to use this model to "reverse engineer" this entire action plan as it unfolds in real time.

To see how this would work, let us add an additional level to the original three-level hierarchical generative model, sketched out in Section 4.1, just below the highest level that represents goals. This new level would represent the set of possible and probable *subgoals* that we believe might be generated by each goal. Each subgoal would generate a range of shorter event clusters, each with different likelihoods. In addition, just like the

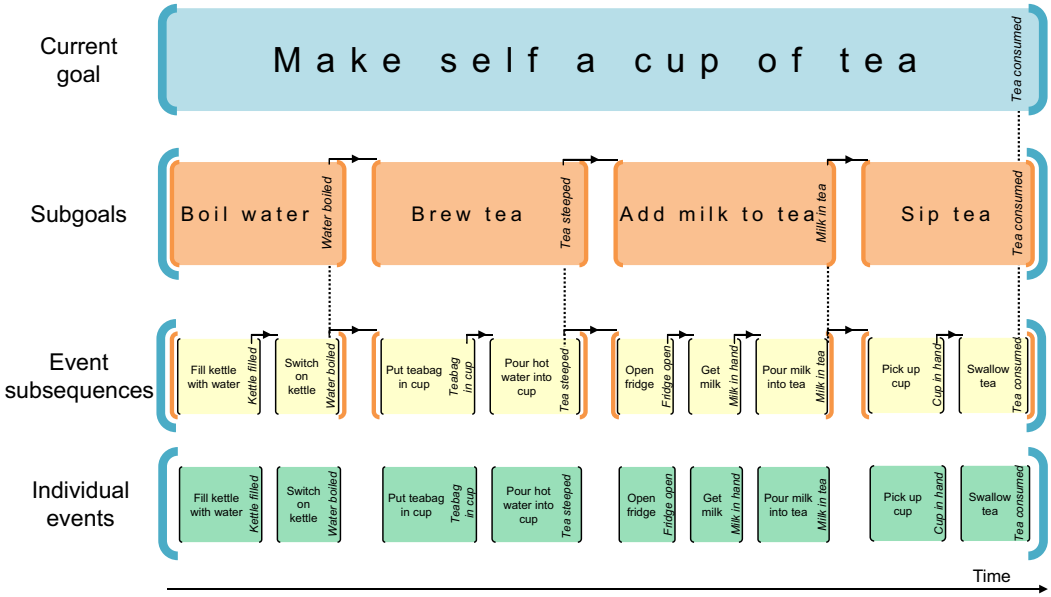

Fig. 3. A schematic of a hierarchical action plan that another agent might use to achieve the goal, {Make self a cup of tea}. Information at successively longer time spans is represented at successively higher levels of the hierarchy. Note that this action plan could, in principle, be expanded to include higher hierarchical levels that would represent longer-term goals, for example, {Make breakfast}, and lower levels that would represent shorter-scale events. For example, <Put teabag in cup> would subsume <Reach for tea canister>, <Open tea canister>, <Reach inside canister>, <Take teabag out of canister>, <Hold hand with teabag over cup>, <Drop teabag in cup> (see also footnote 11).

main goal and each individual event, each subgoal would be represented with its own end state.

This arrangement results in a tree-like hierarchical structure in which any given subgoal would "see" only a subset of event clusters, and the individual events below it. However, the goal at the highest level of the hierarchy would see everything below it. This means that during real-time comprehension, as we progress along a sequence of events and edge closer and closer toward inferring the overarching goal, each subgoal would receive increasingly more top-down pre-activation from this goal. This would offer additional top-down constraints that would further facilitate processing at the levels below. Moreover, the expected rises in uncertainty at the end states of each subgoal would offer us additional flexibility to transition to new subgoals in response to unexpected inputs, with these new subgoals still being informed by top-down activation from the higher-level goal.

To illustrate these points, let us return to the moment in the tea-making sequence when we see that the woman's tea is steeped. Because we are so far along in the sequence, our high certainty over the {Make self a cup of tea} goal will have led to the proactive top-down retrieval of tea-relevant event clusters within working memory at the second level of the hierarchy. However, because [Tea steeped] corresponds to the end state of the

previous subgoal {Brew tea}, a small rise in expected uncertainty about the next subgoal will lead us to disengage slightly from generating strong top-down predictions over single events at the first level of the hierarchy. Therefore, when we encounter the unexpected event, <Open fridge>, our surprise will be slightly more expected and we will be more ready to flexibly update our beliefs and shift to a new subgoal. Moreover, the inference of this new subgoal will be informed by the strong activation of tea-relevant event clusters within working memory—there are only so many subgoals that are both tea-relevant and match <Open fridge>. Therefore, we will be able to infer, with fairly high certainty, that the woman's new subgoal is {Add milk to tea}. This inference will, in turn, lead us to predictively pre-activate the most likely upcoming event in the sequence, <Get milk>.

More generally, the principles illustrated in this simple example offer insights into how dynamic inference would proceed if the hierarchy was expanded still further. Progressively higher levels of the generative model would have increasingly larger (longer) temporal "receptive fields." Lower levels of the hierarchy would process information that changes at faster rates, allowing for rapid and flexible proactive transitions across boundaries at end states (from event to event, and from subgoal to subgoal). Higher levels of the hierarchy would process information that changes more slowly, both because it takes longer to infer new information, and because the information that has been inferred at these higher levels becomes increasingly more resistant to change. Therefore, at successively higher levels of the generative hierarchy, the magnitude of any prediction error and the degree of expected uncertainty required to transition across boundaries would become progressively larger. As a result, during comprehension, unpredicted incoming events will usually be explained at lower rather than higher levels. However, these lower levels will still receive the benefit of top-down pre-activation that is based on whatever has already been inferred at higher levels. Finally, the alignment of certain end states across multiple hierarchical levels means that there are critical points in the information stream where bottom-up information has a short-cut to the very top of the hierarchy, offering us the benefits of a small world network (cf. Watts & Strogatz, 1998; see Minsky, 1975 for early discussion) so that we can proactively attend to inputs in the environment that direct us to whole new branches of discovery.

This perspective highlights the key role of end states—the points of maximal uncertainty at each level of the hierarchy—in proactively directing attention and allowing for flexible shifts between events, subgoals, and goals. This close link between attention and goal end states seems to be fundamental. As discussed by Elsner and Adam (2021, this issue), dynamic eye tracking studies show that, by the end of their first year of life, infants actively attend to the end states of simple familiar events (e.g., Cannon & Woodward, 2012; Falck-Ytter, Gredeback, & von Hofsten, 2006; Kanakogi & Itakura, 2011). They suggest that what may be critical for this type of proactive attention is that the infant believes that the end state of an event corresponds to the end state of another agent's *goal* (Adam & Elsner, 2018; Falck-Ytter et al., 2006; cf. Gergely & Csibra, 2003). This raises the possibility that during development, uncertainty-based tracking of other agents' goals may play a role in building the rich stores of event knowledge that

we rely upon as adults, and that this privileged role of end states in simple events may be rooted in rational Bayesian principles.

5.6. Flexible generative models: The goals of the comprehender

While the type of expanded hierarchical generative model described above can, in principle, allow us to invert the entire action plan of other agents (Schmidt et al., 1978; see Cooper, 2021, this issue; Knott & Takac, 2021, this issue), it is important to recognize that, in practice, this type of "deep deconstruction" of the action plan is often not necessary, and that, ultimately, the structure of the generative models that we engage during comprehension will be determined by our *own* comprehension goals.

It is well established that our mode and depth of comprehension vary considerably, depending on our current "standards of coherence" (cf. van den Broek, Bohne-Gettler, Kendeou, Carlson, & White, 2011). These standards will depend on many factors, including our internal motivation, other tasks we are carrying out, the constraints of our broader environment, and any instructions that we have been given (for discussion in relation to reading comprehension, see Graesser, Singer, & Trabasso, 1994; van den Broek et al., 2011; see Baldwin & Kosie, 2021, this issue for discussion in relation to visual event comprehension). Given that there are metabolic costs of passing information both up and down the cortical hierarchy (Attwell & Laughlin, 2001; Laughlin, de Ruyter van Steveninck, & Anderson, 1998), it would be wasteful to invest resources in pre-activating upcoming information that is irrelevant to our current comprehension goals (see Kuperberg & Jaeger, 2016, p. 13 for discussion; see also Norris, 2006, p. 330). Instead, from a "bounded" rational perspective (see Griffiths, Lieder, & Goodman, 2015; Howes, Lewis, & Vera, 2009; Simon, 1956), it would make more sense to engage a generative model whose depth (the number of hierarchical levels represented) directly reflects the depth of understanding we need at any given time (see Brothers, Wlotko, et al., 2020, for a recent example and discussion; see also Friston et al., 2015, for a more general discussion). For example, if as we watch the woman in her kitchen, our aim is to simply get a general gist of what is going on, then we might engage a shallow generative hierarchy, inferring supergoals like {Make breakfast}, but not more specific goals or subgoals. If, on the other hand, we were being instructed in the art of making tea, then we would be more likely to engage a deeper generative hierarchy that would allow us to infer each individual goal and subgoal (see Hanson & Hirst, 1989, for evidence that our mode of comprehension can influence the hierarchical structure of the information that we encode and store within long-term memory).

Finally, it is important to note that there may be times when we are not interested in the goals of other agents at all, and we approach comprehension with our own set of orthogonal goals and interests. For example, if I came across the woman in the kitchen on YouTube in the midst of a search for ideas about remodeling my own kitchen, then I might gaze intently at my computer, engaged in deep "comprehension." However, at the highest level of my generative model, instead of representing my beliefs about the woman's possible goals, I would instead represent my beliefs about countertop surfaces

and kitchen cabinets. I might additionally be able to ensure that relevant information reaches the top of this generative hierarchy by selectively reducing the variance over relevant perceptual channels at lower levels, thereby increasing the gain on inputs of interest (increasing the magnitude of the bottom-up prediction error that they produce, see Feldman & Friston, 2010). Thus, instead of my attention being proactively captured by my uncertainty in the woman's goals, or retroactively captured by unexpected surprise, it would be (retroactively) captured by inputs that show wood or granite!

This particular example may be too extreme. Human beings have evolved to be interested in other human beings, and so it may be challenging for us to disengage entirely from attending to the goals of other agents. However, it highlights the fact that event comprehension is not always equivalent to event production "in reverse." Instead, the generative models that we engage must be highly dynamic and flexible, allowing us to infer others' goals, but only when this information is relevant to achieving our own goals, as comprehenders.

6. From event comprehension to event production and learning

The premise of the hierarchical generative framework of comprehension described above is that (a) we engage a body of probabilistic knowledge that describes our generalizable assumptions of how and why other agents *produce* events—a *hierarchical generative model*, and (b) we use this model as an inference engine to interpret these events. A basic assumption of this framework is that event comprehension draws upon the same core event representations that we draw upon to carry out sequential action, as discussed in Section 2 (a similar position is taken by Cooper, 2021, this issue; Knott & Takac, 2021, this issue).

Some developmental work supporting this assumption is discussed by Elsner and Adam (2021, this issue). As noted above, young infants actively attend to the end states of simple familiar events (e.g., Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Kanakogi & Itakura, 2011). However, to engage in this type of proactive processing, these infants must have already developed the motor skills required to actually carry out the actions depicted (e.g., Ambrosini et al., 2013; Kanakogi & Itakura, 2011). For example, if an infant has not yet acquired a precision grasp, then while watching a hand precisely grasp a small object, her eyes will track the grasping motion instead of skipping ahead to predict the object (Ambrosini et al., 2013).

6.1. Generative models and the production of sequential action

In addition to simply drawing upon the same underlying event representations as during event comprehension, the production of sequential action may also rely on assembling probabilistic generative models to actively predict upcoming information, just as during comprehension (see Cooper, 2021, this issue; Pezzulo, Rigoli, & Friston, 2018; see also Pickering & Garrod, 2013, for discussion in relation to language production). As

emphasized throughout this review, sequential action necessarily requires us to interact continuously with the environment. Therefore, as in comprehension, predicting upcoming information from the environment before it becomes available from the bottom-up input should increase the speed and efficiency of processing.

A critical role of internal generative models in predicting the consequences of action has long been recognized in the study of simple motor control, where these models are referred to as *forward models*. To effectively carry out even a simple action, such as picking up a cup of hot tea, the motor plan must receive continuous feedback from the perceptual input (e.g., the shape of the cup and the precise temperature of the tea). Waiting for all this perceptual information to become available from the bottom-up input would take too long for it to provide feedback in time to influence the motor plan during real-time action. By actively generating predictions of the perceptual consequence of the motor plan, forward models buy us critical time: Upcoming perceptual information is pre-activated, which means that when it becomes available from the bottom-up input, we only need to compute the difference between what we predicted and what we perceive (just as discussed in Section 4.2 in relation to predictive coding). This "prediction error" can then be used to dynamically update the motor plan "on the fly," ensuring that the movement is smooth and coordinated.

As discussed by Cooper in this issue, similar logic holds at higher levels of the action hierarchy. By engaging a generative model to probabilistically pre-activate the end states of the goals, subgoals, and individual events that we plan to carry out, we gain a head-start. For example, if I am in my kitchen with the subgoal of adding some milk to my tea (because, yes, I do take tea with milk!), then the anticipation of the end state, [Milk in tea], will not only trigger me to carry out the next event, <Open fridge> (cf. Hommel et al., 2001; James, 1890/1981; Lotze, 1852; Prinz, 1987); it may also lead to the pre-activation of the end state of the following event—[Milk in hand]. As a result, as I open the fridge, I know exactly where to look for the milk.

Evidence for this type of proactive prediction during the production of sequential action comes from studies showing that our eyes fixate on objects in our environment just *before* we actually need them during naturalistic goal-directed tasks (Flanagan & Johansson, 2003; Hayhoe & Ballard, 2005), including the task of making tea (see study by Land, Mennie, & Rusted, 1999!). Moreover, within this actively generative framework, when we are very certain about a future goal, not only do we pre-activate the visuo-spatial features of anticipated objects associated with this goal; we also pre-activate the perceptual consequences of our planned actions. For example, as I open the fridge to get the milk for my tea, I would not only pre-activate a representation of the milk in its expected position inside the fridge, but also a representation of my hand actually grasping the milk (see Belardinelli, Lohmann, Farne, & Butz, 2018; Belardinelli, Stepper, & Butz, 2016, for consistent evidence). Therefore, as the fridge opens, not only will my eyes skip ahead to find the milk in its rightful position, but when I come to execute the next action (<Get milk>), I need only to compute the difference between what I predicted and what I perceive and use this prediction error to dynamically update my motor plan as it is executed in real time.

More speculatively, beyond playing a role in predicting upcoming inputs, the generative models we engage during the production of sequential action may also function to track uncertainty in our *own* goals and subgoals, allowing us to "infer" parts of our own action plans based on the environmental input. This idea may at first seem counterintuitive, at least from the perspective of comprehension: If producers already have access to their own goals and subgoals, then why would they need to engage a probabilistic generative model to infer them? However, there may be times when we approach action with a broad goal, but without a full, precise hierarchical plan of exactly what subgoals and events we will execute in what order (see Miller et al., 1960). By tracking our expected certainty about our own goals and subgoals, we would be able to opportunistically exploit the environment to "fill in the gaps," enabling us to achieve our overall objectives (cf. Patalano & Seifert, 1997). For example, I may walk into my kitchen with the overall goal of making myself a cup of tea. However, I am unlikely to have planned exactly when I'll add the milk. Rather, this will depend on what I see. If I notice that a carton of milk is already on the table, I will probably pour some milk into my cup before I boil the kettle. In this case, it is the environmental input itself (the milk on the table), in combination with my prior expected uncertainty over my own subgoals, that leads me to infer the subgoal <Add milk to tea> at that particular time. This newly inferred subgoal, in turn, predictively generates the next events in my action plan.[12]

More generally, tracking expected certainty of our own goals may also allow us to calibrate our "mode of action" to these overarching goals. For example, there are times when I walk into my kitchen, very certain of my goal of making myself a cup of tea. In these cases, if I reach inside the fridge to get the milk and notice some unpredicted mess, I will "downweight" the resulting prediction error and refrain from switching to a new goal. There are other times, however, when I might wander into my kitchen without any strong intention of making tea. I'll see the kettle, start to boil the water, open the fridge to get some milk, and notice the same unpredicted mess. In these situations, I'll more easily update my original high uncertainty tea-making goal and retrieve new event clusters that are relevant to a new goal. Thus, generative models may play a role in ensuring that our actions are calibrated not only to our environment but also to our own intentions, thereby minimizing the chances of errors in production—*pre-emptive error monitoring*.

To sum up, by allowing for active prediction and updating based on the certaining of goals and subgoals, generative models in sequential action may play a similar role to generative models in event comprehension, thereby ensuring that processing is fast, accurate, and flexible. It is, however, also important to recognize that the generative models we engage in comprehension and production will differ in important ways. First, comprehenders will almost always have more uncertainty about the goals of producers than producers have about their own goals. Second, as discussed in Section 5.6, the generative models that we engage during comprehension will only represent possible goals and subgoals of other agents to the degree that these representations align with our own goals.

Third, during comprehension, we must be able to represent schema-relevant knowledge that is different from the knowledge that we would employ if we, ourselves, were fulfilling a particular goal during production. For example, I would not dream of putting sugar

in my own tea, but if I am watching my father making himself a cup of tea, it would be helpful to represent {Add sugar} as a high certainty subgoal that drives the events I observe. The fact that we *know* that other agents draw upon overlapping but distinct sources of schema-relevant knowledge is, of course, crucial for being able to work together collaboratively toward common goals (Tanenhaus, Chambers, & Hanna, 2004), and for communicating with each other through language (see Brown-Schmidt, Yoon, & Ryskin, 2015, for discussion). In these situations, both comprehenders and producers must, together, try to reduce uncertainty and minimize prediction error across *each other's* generative models (Jaeger & Ferreira, 2013).

6.2. Generative models and learning

Although we will never have full access to each others' brains, there is one important way in which we can bring our generative models closer together—we can *adapt* our models by implicitly *learning* from our environment at the same time as we comprehend and act upon it. There is growing evidence that learning is very closely intertwined with both comprehension and production, relying on the same computational algorithms (e.g., Chang, Dell, & Bock, 2006; Dell & Chang, 2014; Elman, 1990; Elman & McRae, 2019; see McRae et al., 2021, this issue for discussion).

Within a probabilistic generative framework, learning, just like comprehension and production, entails using Bayes' Rule to update our beliefs. The key difference is that instead of updating our beliefs about representations that we have already learned (inference), we must update our beliefs about the *parameters* of the generative model itself, which takes place over a longer time scale. To understand this, consider once again the structure of the three-level hierarchical generative model sketched out in Section 4. In describing the parameters that link the third to the second level of the hierarchy (linking goals to schema-relevant clusters), I suggested that, given our lack of knowledge about the woman in the kitchen, these parameters would reflect our beliefs about an "average" woman's schema-relevant knowledge. However, we *know* that the woman in the kitchen is not simply an "average woman"; she, just like every other person on the planet, draws upon her own unique schema-relevant knowledge, including her specific tea-making/drinking habits and preferences. Therefore, as comprehenders, we will always have some uncertainty about these parameters, and it is this *expected uncertainty* that drives us to update our beliefs about these parameters, in parallel with updating our beliefs about the representations of the model itself. Therefore, after watching the woman make herself a few cups of tea, we should be able to implicitly adapt our generative model so that its parameters more accurately describe the specific tea-drinking habits of this particular woman (see Kleinschmidt & Jaeger, 2015, for a generative model of processing and adaptation at a low level of language processing).

Of course, adapting our generative model in this way is just the first step; there is no point in learning that the woman takes her tea with milk if we forget this vital piece of information after a night's sleep! We must be able to encode and consolidate this information within long-term memory. As discussed by several contributors to this Special

Issue (Bilkey & Jensen, 2021, this issue; Shin & DuBrow, 2021, this issue), there is evidence that hippocampal activity at event boundaries plays a role in consolidating information for later recall (see Baldassano et al., 2017; Ben-Yakov & Dudai, 2011).

From a computational perspective, a fundamental question is how we are able to *generalize* from the information that we learn so that we know what schema-relevant clusters are most appropriate to retrieve in future situations. For example, if having learned that the woman in the kitchen takes her tea with milk, when we meet her for tea at another time and place (say, at the Ritz), how do we know whether to offer her milk? Or, suppose that we learn that she is English; do we take our newly encoded woman-in-kitchen-makes-tea-with-milk generative model as the starting point for generalizing about other English tea drinkers? In all these cases, we need to be able to *infer* what kinds of prior experiences are most relevant for the present situation (see Kleinschmidt & Jaeger, 2015, for a detailed discussion of relevant issues). As discussed by Shin and DuBrow (2021, this issue), the Dirichlet process infinite mixture models described above provide insights into how we might be able to learn and extract abstract features that are common to different event clusters, allowing for this type of generalization. These types of models can also explain specific patterns of memory and decision biases (see Franklin et al., 2020).

Taken together, this work underlines the idea that, within this probabilistic generative framework, comprehension, production, learning, and memory are intimately related. This, of course, makes sense. Comprehending, producing, and learning from events are not separate endeavors: All three processes must work alongside one another as we perceive and act upon our environment. By engaging in probabilistic prediction, and tracking both prediction error and uncertainty of our prior beliefs, hierarchical generative models may provide the computational engine that allows us to both exploit and learn from the rich statistical structure of the world around us.

Acknowledgments

This work was funded by the National Institute of Child Health and Human Development (R01 HD082527 to G.R.K.). The review was inspired by reading the other contributions to this Special Issue. I thank Trevor Brothers, Sam Gershman, Nick Franklin, Marcia Kuperberg, and Lin Wang, as well as five reviewers and Martin Butz (Editor of this Special Issue) for their insightful comments on the manuscript. I am also grateful to Mike Jacobson for taking the photos of me making tea in my kitchen that were used in Fig. 2 (as well as for making me many cups of tea as I wrote this review), and to Arim Choi Perrachione for her assistance in manuscript preparation and figure creation.

Notes

1. Thematic roles are neither purely syntactic nor purely semantic. Rather, they lie squarely at the interface between structure and meaning (Levin, 1993), providing a

way to combine semantics and syntax to describe or understand "who does what to whom."

2. As discussed by Unal, Ji, and Papafragou (2021, this issue), *bounded events* are distinguished from *unbounded events* in which the ending is not explicitly coded in the linguistic expression. For example, while the end state, [Tea consumed], is part of the conceptual structure of the event that is conveyed by the sentence, "The woman swallowed the tea," the linguistic expression, "The woman drinks some tea" does not specify the end of the tea drinking. In the real world, however, events usually have an ending—we know that the woman will not drink tea forever.
3. This type of active high-level representation within working memory has been given different names depending on the modality of input. For example, in the reading comprehension literature, it is referred to as the *situation model* (e.g., Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998); in the literature on spoken language comprehension, it is often referred to as the *message* that is being communicated by the producer (e.g., Bock, 1987; Bock & Levelt, 1994; Dell & Brown, 1991), or the *speech act* when the producer's communicative intention is being emphasized (e.g., Levinson, 2013). Even more broadly, it has been referred to as a *mental model* (Johnson-Laird, 1983), or simply as a *contextual representation* (e.g., Kuperberg & Jaeger, 2016).
4. Throughout this review, I use "we" and other terms associated with agency (e.g., "hypothesis" and "belief updating") to describe probabilistic computations as well as neurocognitive mechanisms, adopting Dennett's Intentional stance (Dennett, 1987; see also McGregor, 2017). The relevant processes and neural mechanisms are, however, assumed to be unconscious and implicit.
5. In most models of predictive coding, the information that is passed up from lower to higher levels of the cortical hierarchy is the bottom-up information that *has not* been predicted (positive prediction error) rather than information that has been predicted but not actually observed (negative prediction error, which, in the cortex, is biologically implausible, see Keller & Mrsic-Flogel, 2018; Rao & Ballard, 1999). In network models, positive prediction error can be calculated by computing the element-wise difference between information carried by the bottom-up input and information in the top-down reconstruction at each unit and taking the positive values (e.g., Ballard & Jehee, 2012; Keller & Mrsic-Flogel, 2018), or it can be computed through element-wise division (dividing the input by the prediction at each unit, see Spratling, 2008; Spratling, De Meyer, & Kompass, 2009). It is usually assumed that this prediction error is calculated by "error units" that are distinct from "state units" at each level of the generative hierarchy.
6. As discussed by McRae, Brown, and Elman (2021, this issue), recurrent neural networks are, by definition, dynamic models that represent states that continuously change over time (see Elman, 1990; Jordan, 1986). They have a complex high-dimensional state space and a highly nonlinear transition function. They do not carry out full Bayesian inference. However, they can be viewed as generative and probabilistic, functioning to implicitly estimate the probability distribution of an

upcoming observation in a sequence, given previous observations (e.g., Chung et al., 2015; Rabovsky, Hansen, & McClelland, 2018).

7. Consistent with this view of the amplitude of the N400 reflecting the amount of unpredicted information associated with a new input (the magnitude of the first-level bottom-up prediction error), a plausible incoming event that is unpredicted because it follows a non-constraining prior context produces just as large an N400 as a plausible event that is unpredicted because it violates the constraints of a prior context (Kutas & Hillyard, 1984; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kuperberg, Brothers, & Wlotko, 2020).
8. If the agent does not represent a non-stationary environment, then the detection of unexpected surprise will lead to an overwriting of the original generative model (e.g., Dayan & Yu, 2003), corresponding to the so-called catastrophic inference in connectionist models (McCloskey & Cohen, 1989). If the agent does represent a non-stationary environment, but the incoming event cannot be explained by any existing latent causes (e.g., there are no existing goals or schema-relevant clusters that can explain the input), then the agent will learn a new latent cause to explain the input. This new learning of latent causes can also be modeled by the infinite mixture models described above, providing insights into how, during development, we are able to build our vast body of schema-based knowledge (see Franklin, Norman, Ranganath, Zacks, & Gershman, 2020). In adults, upon encountering an unpredictable event that cannot be explained by the current latent cause, it is usually more parsimonious to switch to a previously learned latent cause (i.e., *retrieve* schema-relevant event clusters from long-term memory) than to learn a new latent cause (learn a new cluster of events). However, even in adulthood, *highly* unexpected surprise may trigger new learning if we encounter highly implausible/impossible events that cannot be explained by existing goals/schema stored within long-term memory (e.g., if, instead of seeing the woman in the kitchen pick up the cup to drink her tea, we see her try to snort the tea through her nose, see Sitnikova, Holcomb, Kiyonaga & Kuperberg, 2008 and footnote 9).
9. This *late frontal positivity* can be contrasted with another late positivity with a *posterior* scalp distribution, otherwise known as the P600. This *late posterior positivity/P600* is produced by events that *cannot* initially be explained by existing goals/schema and are therefore initially interpreted as being highly implausible/impossible (in language comprehension, e.g., Kuperberg, 2007; Kuperberg, Brothers, & Wlotko, 2020; Shetreet, Alexander, Romoli, Chierchia, & Kuperberg, 2019; in visual event comprehension, e.g., Cohn, Jackendoff, Holcomb, & Kuperberg, 2014; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008). This type of *highly* unexpected surprise is thought to trigger a general "orientation" response (Nieuwenhuis, 2011; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Yu & Dayan, 2005), which may be reflected by the well-known P300 ERP component (Donchin & Coles, 1988) to which the *late posterior positivity/P600* is thought to be functionally related (Coulson, King, & Kutas, 1998; Osterhout, Kim, & Kuperberg, 2012; Sassenhagen & Fiebach, 2019; Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). This,

in turn, can lead to a number of downstream effects, including second-pass attempts to check that the input was perceived correctly the first time around (re-analysis, see van de Meerendonk, Kolk, Chwilla, & Vissers, 2009), as well as learning new event clusters (see footnote 8).

10. As noted in Section 4.1.1, *goal end states* are also likely to be represented probabilistically. For example, we may not be certain whether the end state of the woman's {Make self a cup of tea} goal is [Take one sip of tea], with a more gradual transition to another goal, or whether she will take her time drinking the whole cup of tea before she starts a new action sequence. Moreover, although to my mind, this goal implies that the woman will drink the tea once she has made it, this may be unclear to others (e.g., to one of the reviewers of this paper!), and so these observers might hold some degree of belief that the end state is [Tea steeped], leading to a rise in expected uncertainty earlier in the sequence.
11. In principle, it should be possible to recursively "unfold" the entire action hierarchy until we reach the very lowest level that represents the shortest time scale of an event—an "event primitive" (see Miller, Galanter, & Pribram, 1960, for early discussion). In their contribution to this Special Issue, Knott and Takac (2021, this issue) suggest that this primitive is a short-lived deictic routine (cf. Ballard, Hayhoe, Pook, & Rao, 1997), and liken the recursive unrolling of an action hierarchy to the way that syntactic trees can be recursively unfolded to reveal a primitive transitive structure.
12. This perspective links to the literature on model-based reinforcement learning: instead of conceptualizing goal-relevant decision-making as finding the policy that maximizes the magnitude of expected reward, it is conceptualized as maximizing the probability of a potential action-outcome-reward sequence ("planning as inference," see Botvinick & Toussaint, 2012; Solway & Botvinick, 2012).

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, *36*(7), 715–729. https://doi.org/10.1037/0003-066x.36.7.715

Adam, M., & Elsner, B. (2018). Action effects foster 11-month-olds' prediction of action goals for a non-human agent. *Infant Behavior and Development*, *53*, 49–55. https://doi.org/10.1016/j.infbeh.2018.09.002

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Altmann, G. T. M., & Ekves, Z. (2019). Events as intersecting object histories: A new theory of event representation. *Psychological Review*, *126*(6), 817–840. https://doi.org/10.1037/rev0000154

Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLoS ONE*, *8*(7), e67916. https://doi.org/10.1371/journal.pone.0067916

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429. https://doi.org/10.1037/0033-295x.98.3.409

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.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, *21*(10), 1133–1145. https://doi.org/10.1097/00004647-200110000-00001

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113* (3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldi, P., & Itti, L. (2010). Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649–666. https://doi.org/10.1016/j.neunet.2009.12.007

Baldwin, D. A., & Kosie, J. E. (2021). How does the mind render streaming experience as events? *Topics in Cognitive Science*, *13*, 79–105.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–742. https://doi.org/10.1017/S0140525X97001611

Ballard, D. H., & Jehee, J. (2012). Dynamic coding of signed quantities in cortical feedback circuits. *Frontiers in Psychology*, *3*, 254. https://doi.org/10.3389/fpsyg.2012.00254

Barker, R. G., & Wright, H. F. (1954). *Midwest and its children: The psychological ecology of an American town*. Evanston, IL: Row, Peterson and Company.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/Nn1954

Belardinelli, A., Lohmann, J., Farne, A., & Butz, M. V. (2018). Mental space maps into the future. *Cognition*, *176*, 65–73. https://doi.org/10.1016/j.cognition.2018.03.007

Belardinelli, A., Stepper, M. Y., & Butz, M. V. (2016). It's in the eyes: Planning precise manual actions before execution. *Journal of Vision*, *16*(1), 18. https://doi.org/10.1167/16.1.18

Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *Journal of Neuroscience*, *31*(24), 9032–9042. https://doi.org/10.1523/JNEUROSCI.0702-11.2011

Bilkey, D. K., & Jensen, C. (2021). Neural markers of event boundaries. *Topics in Cognitive Science*, *13*, 128–141.

Bock, J. K. (1987). Exploring levels of processing in sentence production. In G. Kempen (Ed.), *Natural language generation* (pp. 351–363). Dordrecht: Martinus Nijhoff.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). London: Academic Press.

Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*(2), 395–429. https://doi.org/10.1037/0033-295X.111.2.395

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*(2), 177–220. https://doi.org/10.1016/0010-0285(79)90009-4

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Brothers, T., Dave, S., Hoversten, L., Traxler, M. J., & Swaab, T. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Brothers, T., Greene, S., & Kuperberg, G. R. (2020). Distinct neural signatures of semantic retrieval and event updating during discourse comprehension. Presented at the 27th Annual Meeting of the Cognitive Neuroscience Society, Boston, MA.

Brothers, T., & Kuperberg, G. R. (in press). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. Journal of Memory and Language.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brothers, T., Wlotko, E. W., Warnke, L., & Kuperberg, G. R. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, *1* (1), 135–160. https://doi.org/10.1162/nol_a_00006

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as contexts in conversation. *Psychology of Learning and Motivation*, *62*, 59–99. https://doi.org/10.1016/bs.plm.2014.09.003

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. https://doi.org/10.1016/j.neunet.2019.05.001

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295x.113.2.234

Chater, N., Crocker, M. W., & Pickering, M. J. (1998). The rational analysis of inquiry: The case of parsing. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 441–468). New York: Oxford University Press.

Chow, W. Y., Lau, E. F., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, *33*(7), 803–828. https://doi.org/10.1080/23273798.2018.1427878

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., & Bengio, Y. (2015). A recurrent latent variable model for sequential data. Presented at the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montréal, Canada.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coderre, E. L., O'Donnell, E., O'Rourke, E., & Cohn, N. (2020). Predictability modulates neurocognitive semantic processing of non-verbal narratives. *Scientific Reports*, *10*(1), 10326. https://doi.org/10.1038/s41598-020-66814-z

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia*, *64*, 63–70. https://doi.org/10.1016/j.neuropsychologia.2014.09.018

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, *65*(1), 1–38. https://doi.org/10.1016/j.cogpsych.2012.01.003

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, *120*(1), 190–229. https://doi.org/10.1037/a0030852

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, UK: Cambridge University Press.

Cooper, R. P. (2021). Action production and event perception as routine sequential behaviours. *Topics in Cognitive Science*, *13*, 63–78.

Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274. https://doi.org/10.1111/cogs.12067

Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297–338. https://doi.org/10.1080/026432900380427

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain responses to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58. https://doi.org/10.1080/016909698386582

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3* (S11), 1218–1223. https://doi.org/10.1038/81504

Dayan, P., & Yu, A. J. (2003). Uncertainty and learning. *IETE Journal of Research*, *49*(2/3), 171–182. https://doi.org/10.1080/03772063.2003.11416335

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the 'model of the listener'. In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Lawrence Erlbaum.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61C*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–427. https://doi.org/10.1017/s0140525x00058027

Dowty, D. R. (1989). On the semantic content of the notion of thematic role. In G. Cherchia, B. Partee, & R. Turner (Eds.), *Properties, types and meaning* (pp. 69–129). Norwell, MA: Kluwer.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Elman, J. L., & McRae, K. (2019). A model of event knowledge. *Psychological Review*, *126*(2), 252–291. https://doi.org/10.1037/rev0000133

Elsner, B., & Adam, M. (2021). Infants' prediction of action-events for human and non-human agents. *Topics in Cognitive Science*, *13*, 45–62.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037//0033-295x.102.2.211

Falck-Ytter, T., Gredeback, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Fillmore, C. J. (1967). The case for case. Presented at the Texas Symposium on Language Universals.

Fillmore, C. J. (2006). Frame semantics. *Cognitive Linguistics: Basic Readings*, *34*, 373–400. https://doi.org/10.1515/9783110199901.373

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured event memory: A neuro-symbolic model of event cognition. *Psychological Review*, *127*(3), 327–361. https://doi.org/10.1037/rev0000177

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: Evidence from brain potentials. *Language, Cognition and Neuroscience*, *31*(9), 1193–1205. https://doi.org/10.1080/23273798.2016.1205202

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/Rstb.2005.1622

Friston, K. J., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*, 151. https://doi.org/10.3389/fpsyg.2012.00151

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/s1364-6613(03)00128-1

Gershman, S. J., Radulescu, A., Norman, K. A., & Niv, Y. (2014). Statistical computations underlying the dynamics of memory updating. *PLoS Computational Biology*, *10*(11), e1003939. https://doi.org/10.1371/journal.pcbi.1003939

Gibson, J. J. (1979). *The ecological approach to visual perception*. New York: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, *20*(1), 41–50. https://doi.org/10.1017/s0140525x97470012

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295x.101.3.371.

Grafman, J. (2002). The structured event complex and the human prefrontal cortex. In D. T. H. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 292–310). Oxford, UK: Oxford University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). New York: Cambridge University Press.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. https://doi.org/10.1111/tops.12142

Gruber, J. S. (1965). *Studies in lexical relations*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. https://doi.org/10.1037/a0030045

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52. https://doi.org/10.1016/j.cognition.2018.02.011

Hanson, C., & Hanson, S. J. (1996). Development of schemata during event parsing: Neisser's perceptual cycle as a recurrent connectionist network. *Journal of Cognitive Neuroscience*, *8*(2), 119–134. https://doi.org/10.1162/jocn.1996.8.2.119

Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General*, *118*(2), 136–147. https://doi.org/10.1037/0096-3445.118.2.136

Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory and Cognition*, *47*(1), 17–32. https://doi.org/10.3758/s13421-018-0847-z

Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, *140*(4), 586–604. https://doi.org/10.1037/a0024310

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The 'wake-sleep' algorithm for unsupervised neural networks. *Science*, *268*(5214), 1158–1161. https://doi.org/10.1126/science.7761831

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878; discussion 878–937. https://doi.org/10.1017/s0140525x01000103

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751. https://doi.org/10.1037/a0017187

Jackendoff, R. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, *18*(3), 369–411.

Jaeger, T. F., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*(4), 359–360. https://doi.org/10.1017/S0140525X12002762.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Macmillan/Harvard University Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.

Jordan, M. I. (1986). Attractor dynamics and parallellism in a connectionist sequential machine. In Clifton, C (Ed.), *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 531–546). Hillsdale, NJ: Erlbaum.

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*, 341. https://doi.org/10.1038/ncomms1342

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kleinschmidt, D. F., & Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Knoblock, C. A. (1992). An analysis of ABSTRIPS. In J. Hendler (Ed.), *Artificial Intelligence Planning Systems: Proceedings of the First Conference (AIPS 92)* (pp. 126–135). San Mateo, CA: Morgan Kaufman.

Knott, A., & Takac, M. (2021). Roles for event representations in sensorimotor experience, memory formation and language processing. *Topics in Cognitive Science*, *13*, 187–205.

Kosie, J. E., & Baldwin, D. (2019). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles and Implications*, *4*(1), 8. https://doi.org/10.1186/s41235-019-0157-4

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616. https://doi.org/10.1080/23273798.2015.1130233

Kuperberg, G. R., Brothers, T., & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246. https://doi.org/10.1162/jocn.2010.21452

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72–79. https://doi.org/10.1016/j.tics.2007.11.004

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Laughlin, S. B., de Ruyter van Steveninck, R. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, *1*(1), 36–41. https://doi.org/10.1038/236

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/josaa.20.001434

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Levinson, S. C. (2013). Action formation and ascription. In T. Stivers & J. Sidnell (Eds.), *The handbook of conversation analysis* (pp. 103–130). Malden, MA: Wiley-Blackwell.

Lotze, H. (1852). *Medicinische Psychologie oder Physiologie der Seele*. Leipzig: Weidmann.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*(4), 590–604. https://doi.org/10.1162/neco.1992.4.4.590

MacKay, D. J. C. (2003). Model comparison and Occam's razor. In *Information theory, inference, and learning algorithms*, pp. 343–356. Cambridge, UK: Cambridge University Press.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 24, pp. 109–165). San Diego, CA: Academic Press.

McGregor, S. (2017). The Bayesian stance: Equations for 'as-if' sensorimotor agency. *Adaptive Behavior*, *25* (2), 72–82. https://doi.org/10.1177/1059712317700501

McRae, K., Brown, K. S., & Elman, J. L. (2021). Prediction-based learning and processing of event knowledge. *Topics in Cognitive Science*, *13*, 206–233.

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*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–281). New York: McGraw-Hill.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–1046. https://doi.org/10.1038/Nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/Jneurosci.0822-10.2010

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: W.H. Freeman.

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*(4), 979–999. https://doi.org/10.1037/0033-295X.112.4.979

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28–38. https://doi.org/10.1037/h0035584

Newtson, D. (1976). Foundations of attribution: The perception of ongoing behavior. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1, pp. 223–248). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436–450. https://doi.org/10.1016/0022-1031(76)90076-7

Nieuwenhuis, S. (2011). Learning, the P3, and the locus coeruleus-norepinephrine system. In R. Mars, J. Sallet, M. Rushworth, & N. Yeung (Eds.), *Neural basis of motivational and cognitive control* (pp. 209–222). Cambridge, MA: MIT Press.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciouness and self-regulation. Vol. 4: Advances in research and theory* (pp. 1–18). New York: Plenum Press.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

O'Reilly, J. X. (2013). Making predictions in a changing world-inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*, 105. https://doi.org/10.3389/fnins.2013.00105

Osterhout, L., Kim, A., & Kuperberg, G. R. (2012). The neurobiology of sentence comprehension. In M. Spivey, M. Joannisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 365–389). Cambridge, UK: Cambridge University Press.

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*(4), 426–448. https://doi.org/10.1016/j.jml.2012.07.003

Parsons, T. (1990). *Events in the semantics of English: A study in subatomic semantics*. Cambridge, MA: MIT Press.

Patalano, A. L., & Seifert, C. M. (1997). Opportunistic planning: Being reminded of pending goals. *Cognitive Psychology*, *34*(1), 1–36. https://doi.org/10.1006/cogp.1997.0655

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552. https://doi.org/10.1037//0033-295x.87.6.532

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In Waltz, D (Ed.), *Proceedings of the American Association for Artificial Intelligence National Conference on AI* (pp. 133–136). Pittsburgh, PA.

Pearl, J. (1987). Distributed revision of composite beliefs. *Artificial Intelligence*, *33*(2), 173–215. https://doi.org/10.1016/0004-3702(87)90034-8

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, *22*(4), 294–306. https://doi.org/10.1016/j.tics.2018.01.009

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347. https://doi.org/10.1017/S0140525X12001495

Prinz, W. (1987). Ideo-motor action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 47–76). London: Routledge.

Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, *3*, 228. https://doi.org/10.3389/fpsyg.2012.00228

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, *2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(6), 608–620. https://doi.org/10.1002/wcs.133

Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision Research*, *39*(11), 1963–1989. https://doi.org/10.1016/s0042-6989(98)00279-x

Rao, R. P. N., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, *9*(4), 721–763. https://doi.org/10.1162/neco.1997.9.4.721

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. https://doi.org/10.1080/15326900701399913

Rumelhart, D. E. (1975). Notes on a schema for stories. *Representation and Understanding: Studies in Cognitive Science*, *211*, 236.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. https://doi.org/10.1037/a0020511

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425–436. https://doi.org/10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

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

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schmidt, C. F. (1976). Understanding human action: Recognizing the plans and motives of other persons. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 47–67). Hillsdale, NJ: Lawrence Erlbaum.

Schmidt, C. F., Sridharan, N. S., & Goodson, J. L. (1978). The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence*, *11*(1–2), 45–83. https://doi.org/10.1016/0004-3702(78)90012-7

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770

Shetreet, E., Alexander, E. J., Romoli, J., Chierchia, G., & Kuperberg, G. R. (2019). What we know about knowing: Presuppositions generated by factive verbs influence downstream neural processing. *Cognition*, *184*, 96–106. https://doi.org/10.1016/j.cognition.2018.11.012

Shin, Y. S., & DuBrow, S. (2021). Structuring memory through inference-based event segmentation. *Topics in Cognitive Science*, *13*, 106–127.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. https://doi.org/10.1162/jocn.2008.20143

Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychological Review*, *119*(1), 120–154. https://doi.org/10.1037/a0026435

Spratling, M. W. (2008). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2016a). A neural implementation of Bayesian inference based on predictive coding. *Connection Science*, *28*(4), 346–383. https://doi.org/10.1080/09540091.2016.1243655

Spratling, M. W. (2016b). Predictive coding as a model of cognition. *Cognitive Processing*, *17*(3), 279–305. https://doi.org/10.1007/s10339-016-0765-6

Spratling, M. W., De Meyer, K., & Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational Intelligence and Neuroscience*, *2009*, 381457. https://doi.org/10.1155/2009/381457

Tanenhaus, M. K., Chambers, C. G., & Hanna, J. E. (2004). Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 279–317). New York: Psychology Press.

Tauber, S., Navarro, D., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, *124*(4), 410–441. https://doi.org/10.1037/rev0000052

Unal, E., Ji, Y., & Papafragou, A. (2021). From event representation to linguistic meaning. *Topics in Cognitive Science*, *13*, 224–242.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. https://doi.org/10.1162/089892999563724

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*(5), 1211–1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x

van den Broek, P., Bohn-Gettler, C. M., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. Crudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–139). Charlotte, NC: IAP Information Age.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, *66*(2), 143. https://doi.org/10.2307/2182371

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*(3), 363–375. https://doi.org/10.1016/s0926-6410(01)00129-x

Yang, S. C. H., Wolpert, D. M., & Lengyel, M. (2016). Theoretical perspectives on active sensing. *Current Opinion in Behavioral Sciences*, *11*, 100–108. https://doi.org/10.1016/j.cobeha.2016.06.009

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037/0096-3445.130.1.29

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162