Journal of Experimental Psychology:
Human Perception and Performance
2016, Vol. 42, No. 2, 235–246

0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000082

Default Processing of Event Sequences

Alicia Hymel, Daniel T. Levin, and Lewis J. Baker
Vanderbilt University

In a wide range of circumstances, it is important to perceive and represent the sequence of events. For example, sequence perception is necessary to learn statistical contingencies between events, and to generate predictions about events when segmenting actions. However, viewer's awareness of event sequence is rarely tested, and at least some means of encoding event sequence are likely to be resource-intensive. Therefore, previous research may have overestimated the degree to which viewers are aware of specific event sequences. In the experiments reported here, we tested viewers' ability to detect anomalies during visual event sequences. Participants viewed videos containing events that either did or did not contain an out-of-order action. Participants were unable to consistently detect the misordered events, and performance on the task decreased significantly to very low levels when performing a secondary task. In addition, participants almost never detected misorderings in an incidental version of the task, and performance increased when videos ended immediately after the misordering, We argue that these results demonstrate that viewers can effectively perceive the elements of events, but do not consistently test their expectations about the specific sequence of natural events unless bidden to do so by task-specific demands.

Keywords: event perception, visual cognition, sequence perception

A tradition of research exploring visual awareness suggests that some forms of visual processing are based upon a sparse representation of visual properties that is limited to a small subset of critical features that are the focus of visual attention (Ballard, Hayhoe, & Pelz, 1995; Hayhoe, 2000; Rensink, 2000; Simons & Levin, 1998). On this view, people's visual system sometimes relies upon the external world as an "outside memory" that can serve as a relatively stable informational resource that can make internal representations redundant (O'Regan, 1992). Although it is supported by a range of observations, this sparse representation view requires qualification for two primary reasons. First, any given measure of visual representations cannot be exhaustive, leaving open the possibility that other measures may reveal previously untapped representations (Simons, 2000). Second, there is a vast range of circumstances, tasks, and stimuli, and although some of these must rely upon internal models of the world, it remains important to consider the possibility that an outside memory may be sufficient for others. So, an effective understanding of visual intelligence may require a description of how the balance between internal representations and external informational resources is managed in support of different visual tasks (see, e.g., Triesch, Ballard, Hayhoe, & Sullivan, 2003). The goal of the current article is to extend exploration of this critical balance from the perception of scenes and simple brief events to the perception of longer-lasting meaningful events. In particular, we ask whether viewers represent the sequence of actions that make up familiar meaningful events, and can therefore easily detect misorderings of action sequences (e.g., action sequence violations in which one action is seen after another action that would typically precede it; e.g., an actor might be seen using a tool and then grabbing it.)

Misapprehensions in the sequence of brief auditory events have been known for some time (for a review, see Sinico, 1999), and a few studies have tested participants' ability to detect temporal violations in sequences of event labels and still images (Raisig, Welke, Hagendorf, & van der Meer, 2007, 2010; Ruby, Sirigu, & Decety, 2002), or have assessed the effects of matches in image sequence between learning and test (Aginsky, Harris, Rensink, & Beusmans, 1997). However, we know of no research that has tested viewers' ability to detect misorderings in the sequence of dynamic familiar visual events. Not only is this an important and underexplored question in itself (Koenderink, 2012), but theories of event perception and sequence learning suggest that at least some of the encoding and comparison processes necessary to detect misordered action sequences occur as a default in support of event perception and understanding. The most important example is the class of theories hypothesizing predictive mechanisms for perceiving and understanding events that have long been discussed both in the vision science literature, and in the broader cognition literature. Both of these literatures have assumed that such mechanisms are crucial not only to facilitate basic perceptual extrapolation, but also at a more conceptual level to create models of the near future that can be checked against newly identified events. For example, Zacks's event segmentation theory hypothesizes that individuals segment events by engaging in continuous prediction, contingent upon sequential regularities, with event boundaries

This article was published Online First September 7, 2015.

Alicia Hymel, Daniel T. Levin, and Lewis J. Baker, Department of Psychology and Human Development, Vanderbilt University.

Alicia Hymel is now at the Peabody Research Institute, Vanderbilt University.

This material is based on work supported by the National Science Foundation Grant 0826701.

Correspondence concerning this article should be addressed to Daniel T. Levin, Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203. E-mail: daniel.t.levin@vanderbilt.edu

occurring when these predictions fail (Zacks, Speer, Swallow, Braver, & Reynolds, 2007).

However, research exploring the cognitions that organize event perception demonstrates that at least some of these predictions are resource intensive, suggesting that there are important limits to the generation of the online predictions that serve as the basis for event perception. In this article, we report experiments testing these limits, and demonstrate that it is quite difficult to detect misorderings in event sequences, even when participants can effectively understand the events and detect brief elements of the events. We argue that these results place important limits on the mechanisms responsible for event perception, as participants do not seem aware of the specific sequence of subevents that make up larger events.

Perceptual and Conceptual Predictions in Event Perception

Researchers have explored different ways in which prediction may be used in perceiving and understanding events. At a basic perceptual level, research demonstrating the flash-lag phenomenon suggests that the location of a briefly seen object can be misperceived as having progressed along its path of motion (Nijhawan, 1994), and similar effects have been observed when assessing immediate memory for a moving object. For example, in the first of a long series of studies exploring the phenomenon of representational momentum, Freyd and Finke (1984) showed participants a stationary rectangular stimulus that rotated in an implied direction of motion across multiple presentations. When memory for the final orientation of the stimulus was tested, participants had difficulty rejecting a distractor stimulus that was rotated further in the implied direction of motion than had been previously shown. Similar predictive phenomena can be observed when both adult and infant participants reach to the future location of a moving object (see, e.g., Kerzel & Gegenfurtner, 2003; Von Hofsten, Vishton, Spelke, Feng, & Rosander, 1998).

However, prediction is not necessarily limited to simple forward projections of motion. There is some evidence to suggest that knowledge can affect these lower-level processes (for a review, see Hubbard, 2005). For example, Vinson and Reed (2002) demonstrated that the size of displacement in representational momentum can be affected by prior knowledge of the stimulus object's typical motion. An object that typically moves great distances, such as a rocket, elicited more displacement along its implicit direction of motion than an object that is typically stationary, such as a building. In addition, forward extrapolation is not limited to the trajectory of simple objects. Thornton and Hayes (2004) showed participants brief film clips of everyday scenes and found that participants falsely selected stills taken from a slightly later point than the end of the clips as being the ending they had actually seen.

The use of predictions in both perception and action can be even more conceptual. Theory of mind is the process by which individuals reason about the mental state of others, including their intentions and goals (Premack & Woodruff, 1978), and it is often described as being, at its core, a process devoted to predicting the behavior of others (see, e.g., Baron-Cohen, Leslie, & Frith, 1985). However, although the prediction occurring in representational momentum and even reaching appears to be automatic and effortless, this is not necessarily true for predictions based on more complex events, or for events involving beliefs about others' mental states. Developmental work suggests that children find it easier to understand events that do not require prediction compared with very similar events that do require prediction (for a review, see Keen, 2003), and predicting behavior based on false beliefs appears to require significant mental effort (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006), which can create situations where even adult participants fail to account for a partner's beliefs (Barr & Keysar, 2005). Even in cases where participants do apparently fill in a meaningful unseen event, the inference could just as easily be a postdiction based on an analysis of events following the critical event as a prediction based on events that preceded it (Strickland & Keil, 2011). Finally, models of text inference suggest that predictive inferences do not occur by default (Graesser, Singer, & Trabasso, 1994), and even in situations where naturalistic visual events provide support for predictive inferences, these occur only infrequently, which can be interpreted as inconsistent with "promiscuous prediction" models (Magliano, Dijkstra, & Zwaan, 1996).

Even in cases of apparently lower-level perceptual prediction, it is not clear that predictive mechanisms are always necessary or particularly efficient. For example, the well-known linear optical trajectory model of catching relies upon an explicitly nonpredictive control mechanism for arriving at the correct location to catch a flying ball (McBeath, Shaffer, & Kaiser, 1995). In addition, evidence from phenomena such as color-phi variants of apparent motion (Kolers & von Grunau, 1976), and research demonstrating that the paths subsequent to inducing motion modulate the flash-lag phenomenon (Eagleman & Sejnowski, 2000) suggest that both apparent motion and flash-lag can involve postdictive reconstruction of recent perceptual experience as well as prediction.

Previous research provides evidence that perceptual and conceptual predictions are both useful processes in perceiving and understanding the motion of objects and the behaviors of agents. However, many of these more complex predictions, especially of social behavior, appear to require significant mental effort and therefore are unlikely to be consistently applied as the basis for an abstract conceptual online understanding of events. It is therefore particularly interesting to consider the role of prediction in organizing the more basic sequence of subevents that serve as the basis for this higher-level understanding. This form of organization is more complex than simple forward projections of object motion but less complex than the deep cognitive analysis implied in prediction derived from theory of mind.

Research exploring the basic organization of events has focused on prediction in two basic ways. First, research on statistical learning tests participants' ability to learn contingencies between events derived from experience with event sequences (see, e.g., Baldwin, Andersson, Saffran, & Meyer, 2008), and there is active debate over whether this form of learning requires prediction from one event to the next, or whether more simple associative or memory-based mechanisms are sufficient to learn these contingencies (for a review, see Dale, Duran, & Morehead, 2012). Second, research exploring how viewers segment events has been guided by an explicitly predictive model (Zacks, Tversky, & Iyer, 2001).

Event segmentation theory (Zacks et al., 2007) provides a broad framework for understanding the segmentation process, and a central part of the theory is the generation of predictions about upcoming events. These predictions represent the near future, and they can be as simple as the forward displacement of an object's motion or as complex as predicting a person's actions based on their presumed goals. According to event segmentation theory, predictions are continuously generated after multisensory input is transformed into semantically rich mental representations, a process that is guided by stable event representations held in working memory (known as event models and event schemata). These models contain prior knowledge about events, including conceptual information about the likely content of an event, such as "which patterns of activity are likely to follow a given pattern, and information about actors' goals" (Zacks et al., 2007, p. 275). An error detection mechanism compares these predictions with reality. If the current event models are an appropriate representation of reality, then little prediction error should occur. However, if the event models are not suitable, the predictions will not be accurate, and prediction error is "fed back to update working memory" (Zacks, Kurby, Eisenberg, & Haroutunian, 2011, p. 4057). The resulting increase in prediction error causes the updating of the event model, which is perceived as an event boundary.

A key feature event segmentation theory is that events are processed at multiple levels of abstraction simultaneously, although people may selectively attend to only one level (Zacks et al., 2007). So, the event "brushing teeth" can be contained within the event "getting ready for work," and it can contain the event "putting toothpaste on toothbrush." Accordingly, events are, to some degree, identified at all levels and predictions about subsequent events are also generated at all of these levels. These levels of abstraction could, in principle, be extended quite a bit (Schwann & Garsoffky, 2008). For example, the above events could be contained with events as broad as "not getting fired for being late" and "earning a living," and it could contain "preparing a saccade to the toothpaste." However, Zacks et al. (2007) proposed that timescales ranging from seconds to tens of minutes are an appropriate range of levels of analysis.

It is worth noting that these mismatch-induced prediction errors, or their immediate sequelae, should produce conscious experiences if they are to trigger the construction of a new event model that can create new predictions about an agent's goals. This level of awareness is, in part, suggested by the fact that the primary measure of this signal is an explicit decision to segment events (Newtson & Engquist, 1976; Zacks et al., 2001). Properties of the neural system charged with generating prediction errors reinforce the impression that the necessary processing requires a high degree of sophistication. Zacks et al. (2011) proposed a midbrain phasic dopamine system composed of brain areas responsive to the expectedness of rewards as the generator of prediction errors, a form of processing that requires a comparison between identified events and consideration of goals. If this consideration of goals is to include the goals of observed agents (as implied by the nature of most stimuli used to test event perception), the theory of mind research described above (Apperly et al., 2006; Barr & Keysar, 2005) makes clear that this form of processing can require a level of capacity-absorbing awareness that perceivers often conserve (to the point of failing to account for important factors affecting an agent's goals). Accordingly, if continuously generated predictions are to inform a relatively complete analysis of events, either prediction errors themselves, or the analysis of events that follows from them would seem to require the kind of schema-matching and working memory operations of which viewers should be aware.

If models of event perception imply that viewers represent sequence, some models of narrative comprehension are even more explicit in arguing for a default sequence-representation process. For example, Claus and Kelter (2006) argue that "mental representation of time course of a dynamic situation is a prerequisite for understanding" (p. 1042), and Raisig et al. (2010) argue that it is "crucial that violations of the temporal order [of event sequences] are detected in order to adjust behavior or re-analyze the situation" (p. 1). Of course, in at least some situations this is clearly true - understandings of physical causality and higher-level scripts do sometimes critically hinge on the ability to discriminate one sequence from another. However, it may be an overgeneralization to suppose that sequence representation is necessary in more than a subset of crucial instances, and even evidence presented by the above researchers in support of this process hints at this possibility. For example, Raisig et al. (2010) present subjects with triads of correctly or incorrectly ordered event labels (sequentially for 1,500 ms each) and observed miss rates of up to 14%–15%. Similarly, Ruby et al. (2002) found that participants erred in about 10% of trials and took 3,800 to 3,910 ms to detect misorderings in event sequences depicted in three simultaneously presented drawings. Although this cannot be considered extremely poor performance, it is not characteristic of the near zero error rates and the rapid RTs one might expect for an automatic default process.

In summary, little research directly assesses viewers' awareness of the sequence of familiar events, but research and theory underlying event and narrative perception strongly suggests that some awareness of these sequences is necessary to correctly segment events. Research exploring sequence learning suggests something similar, although this work is more compatible with the hypothesis that sequence processing results in a form of learning that may occur outside of awareness. However, a predictive model that relies upon deep processing and knowledge-guided comparison may be too resource-intensive for efficient perception of at least some meaningful events. Accordingly, an alternative process might forego an intensive default analysis of the specific sequence of events, and to instead assume that event sequence is constrained by simple real-world object-to-object interactions and therefore does not need to be internalized. On such a view, perceivers might be able to effectively understand events even if they cannot, or do not fully perceive the actual sequence of those events. To investigate this possibility, we created live-action movies, some of which contained an out-of-order event. If an individual watched an event that did not conform to standard sequential structure, then the momentary predictions made about that event should be incorrect and produce prediction error. If this type of continuous prediction is crucial to understanding ongoing events, perceivers should be aware of this violation of their prediction, and the misordering should be detected.

Experiment 1

In Experiment 1, we tested participants' ability to detect event misorderings in videos of event sequences with and without interference from a secondary task. If event sequence predictions occur by default, and if predictive failures automatically produce a signal that could induce event segmentation, then one would expect that secondary task interference would not lessen misordering detection.

Method

In these videos, we created misorderings by filming action sequences from multiple points of view similar to those that a filmmaker might select when showing actions in a series of medium shots (e.g., depicting a person sitting at a table engaging in a simple activity) and close-ups (e.g., showing their hands using a tool). This allowed us to reorder actions by reordering critical shots. For example, one of the 12 videos in Experiment 1 contained a misordered action sequence in which an actor was first shown using a screwdriver to work on a telephone, then shown picking up the screwdriver he had already been using, and finally shown opening the phone (see Figure 1). This can be compared with the correct action sequence of picking up the screwdriver from the table, using it on the phone, and finally opening the phone.

Before continuing, it is important to preview key features of our stimuli and to explicitly consider what misordering-detection failures can tell us about event perception. First, our stimuli contain two important events that may conflict with online predictions: an ellipsis and a return to an action that should already be complete. So, in the reversed screwdriver video, the first potential anomaly is that actor can be seen using the screwdriver without being seen to pick it up. This kind of elliptical edit is quite common in cinema, and research indicates that events are perceived and remembered similarly even when relatively unimportant subevents are dropped from a sequence (Schwan & Garsoffky, 2004); indeed, our participants *never* reported these ellipses. Although this may have important implications for the nature of prediction in event perception, our primary focus here is on the return to an event that should already be completed (e.g., seeing the actor using the screwdriver and then picking it up). This event is not part of common sequences, and is likely to be much more clearly anomalous. As confirmation of this, even in situations where viewers were able to detect a misordering, they almost never false alarmed to any other shot-to-shot edit in the sequences, and false alarm rates were very low for clips that did not contain misorderings (mean of 6.7% for no-interference videos).

A second key point to make about misordering detection is that we are testing the degree to which viewers can detect a misordering of subevents within a single larger event. This is the level of analysis most applicable to event segmentation theory, and it most clearly tests participants' perception of event sequence per se. Our aim was to test the degree to which participants perceive a sequence of temporally ordered events that all meet one clear goal, as an analysis of these sorts of subevents is hypothesized to generate predictive mismatches that may indicate that a new goal has become visible. Clearly, it might also have been possible to test reversal detection when actions instantiating higher level goals have been reordered (e.g., showing someone washing the dishes and then eating) but these larger goal sequences are much less constrained (e.g., it is reasonable that someone might need to wash dishes before eating as opposed to washing dishes after they have used them), and the misordering would involve mismatches between semantic event categories (e.g., disrupting a dishwashing event with a categorically different eating event) in addition to misordering of event sequences per se.

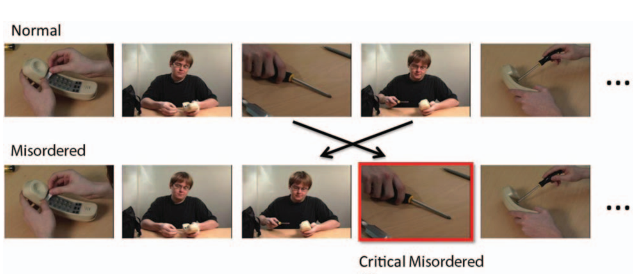

Figure 1. Examples of normal and misordered sequences. Images used with permission of the depicted actor. See the online article for the color version of this figure.

The stimuli we have used to test for misordering detection were designed to make each individual subevent clearly perceptible while preventing simple perceptual cues from making misorderings evident. Therefore, we inserted brief 2-frame (67ms) blank screens between each shot in the video to prevent sudden backward jumps in object position that might otherwise occur in the misordered videos. In a sense, these blanks are analogous to the flickers in change blindness research (Rensink, O'Regan, & Clark, 1997) because they prevent simple exogenous cues from attracting attention to the misordering. For example, if we had cut directly from the second to the third shots of the misordered sequence in Figure 1, the actor's hand would suddenly have jumped forward as the screwdriver popped into it. We also wanted to ensure that the sequences were fully perceptible and understandable while minimizing participants' ability to engage in nondefault elaboration on the videos. Therefore, we edited the videos at a relatively quick pace and sometimes speeded individual shots. This allowed us to test whether reversal detection is central to basic event perception, as opposed to elaboration that may occur as viewers begin to think deeply about a slowly paced event. Because these videos are somewhat unnatural in appearance, we verified that participants could perceive the critical misordered subevents and could effectively understand the sequences as a whole in Experiments 3 and 4.

Participants. Fifteen Vanderbilt University undergraduates (13 female; *M* age = 18.80 years, *SD* = 0.68) participated in exchange for course credit. All participants in this report completed informed consent at the start of the experiment, and all experiments were approved by Vanderbilt University's institutional review board.

Stimuli and procedure. All videos were played at a resolution of 740 × 480 on a 17-inch screen with a 1280 ×1040 resolution. Videos were surrounded by a black border that filled the remainder of the screen. All videos were full color Apple QuickTime files compressed using DV/DVCPRO at a frame rate of 29.97 per second and with no audio.

Each participant watched 12 video sequences consisting of an actor performing a familiar task (e.g., making a cup of coffee, using a copy machine). Each sequence was filmed using multiple camera angles, and shots were a mix of medium shots covering all or most of an actor, and close-ups centered on the most relevant actions. Each shot within a video sequence was separated by 67 ms (two frames) of blank screen. These cuts separated the sequences into a set of discernible subactions (e.g., stirring a cup of coffee, lifting the lid of a copy machine). Additionally, some clips in these videos were slightly sped up to keep clip length brief. One representative example of this was a shot of a woman walking across a room in a video about sharpening a pencil. For all participants, half of the videos showed the events in the correct order, and the other half contained a misordering in which the order of two actions was reversed. Two versions of each of the videos were created—one showing the sequence of actions in the correct order and another that contained the out-of-order sequence. Both the identity of the videos containing an out-of-order action and the order of video presentation were counterbalanced across participants. The average length of the videos was 9,583 ms, the average length of the misordered shots was 533 ms (range = 300–1,066 ms), and there was an average of 11.25 shots within the videos.

Participants were told that they would see a countdown followed by a sequence of events depicting an everyday activity, broken up into shots separated by blank intervals. They were also told that for some of the sequences all of the shots would be in the correct order, but in others a pair of shots would be out of order, and that their task was to detect the misorderings. Participants were instructed to mark on an answer sheet whether or not they saw a misordering after watching each video. Additionally, for half of the trials, participants were told they would have to count backward from a given number by three. Before these trials, this number was displayed on the screen for 5,000 ms along with a brief reiteration of the counting instructions. Participants were told to count out loud during the entire length of the video. The participants' starting numbers were randomly generated and did not differ between participants, although different numbers were assigned to the same videos across participants. The presence of this secondary task was balanced such that four equally frequent trial types existed: misordering and a counting task, correct order and counting task, misordering and no counting task, correct order and no counting task. Across participants, each of the 12 videos represented each trial type approximately equally often.

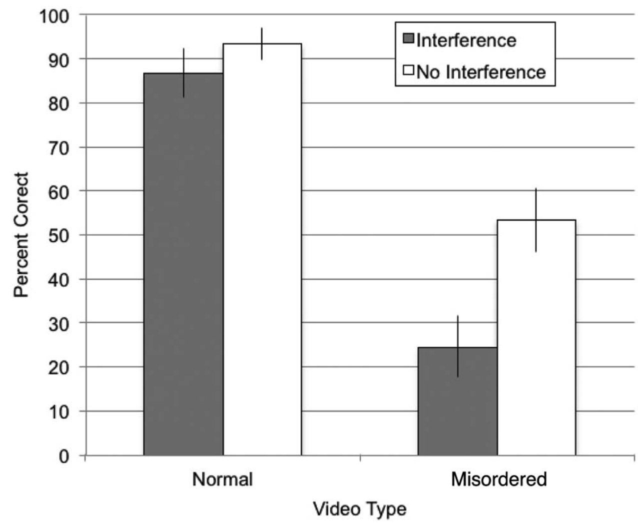

Figure 2. Reversal detection with and without interference task in Experiment 1.

Results

Participants failed to detect a substantial proportion of misorderings, and detected significantly more misorderings on the no-interference trials (53.3%) than on the interference trials (24.4%), $t(14) = 2.83$, $p = .01$. There was no difference in accuracy on the no-misordering trials (93.3% correct rejects for no-interference trials, and 86.7% correct rejects on the interference trials, $t < 1$; see Figure 2). This produced a d' of 1.58 ($c = .708$) for the no-interference trials, and a d' of .418 ($c = .902$) for the interference trials. Overall, performance on the interference trials was not significantly above chance (56% correct), $t(14) = 1.234$, $p = .23$, and was significantly less than performance on the no-interference trials (73% correct), $t(14) = 2.874$, $p = .012$.

Discussion

Experiment 1 demonstrates that detection of sequence reversals was difficult with full attention, and essentially impossible with divided attention. Although these data suggest that reversal detection requires attention and therefore may not be a default process during normal event viewing, it is possible that participants would see a similar proportion of reversals during the normal course of event perception if normal event perception does involve devotion of limited capacity resources to sequence encoding. On this hypothesis, our secondary task interfered with a default encoding process that viewers would have normally applied when processing the events. In Experiment 2, we tested this hypothesis by changing reversal detection into an incidental task. We asked participants to simply pay close attention to the events in a single reversal-containing video in anticipation of answering unspecified questions about them. Participants then viewed the video and were asked whether they detected the reversal.

Experiment 2

Method

Participants. Twenty-eight participants (20 female; *M* age = 20.89 years, *SD* = 4.56) were recruited from Vanderbilt University and the surrounding community. Participants were monetarily compensated. One participant was given the incorrect form and was not included in the analysis, leaving a total of 27.

Procedure. All videos were played at a resolution of 740 × 480 on a 13-inch laptop screen with a 1280 × 800 resolution using Final Cut Pro. Participants viewed the same videos used in Experiment 1, but only videos with an out-of-order action were used.

All participants were told they would be watching a short video of an actor performing an everyday activity, and they would answer questions about the video after watching. In contrast to Experiment 1, they were not informed of the presence of the misordering, and were not told to detect any particular target. Every participant only watched one of the original 12 misordered videos. Most videos were viewed by two to three participants although two were seen by four participants and two were seen by only one because of a sequence error. Immediately after watching the video, participants were asked a series of questions to determine whether they noticed the out-of-order event. Participants first answered whether they noticed anything unusual about the video. Next, they answered whether they noticed anything unusual about the order of events in the video. Finally, they answered whether they noticed one of the actions in the video was out of order.

Results and Discussion

None of the 27 participants reported noticing anything unusual about the order of events in the videos or that one of the actions in the video was out of order. Although some participants reported noticing something usual about the video, in follow-up questioning, this usually related to the video's editing (e.g., the presence of the black screen between individual shots). In one case, one participant thought that something repeated, and that the actor "might have turned the flashlight on before replacing the cap" (in fact, the misordering in this video was that the actor could be seen putting the top back on the flashlight after he could be seen screwing it on).

This experiment provides further evidence that misordering detection does not occur as a default during event viewing. However, this claim requires further testing before it can be accepted, in part because the videos we used have a number of features that may have obscured the basic subevents themselves. Although all of the subevents were visible for long enough for basic scene perception and object identification to take place, the videos were edited to make the events occur in rapid succession, and the blank screens separating the clips may have interfered with perception of the clips they separated. It is possible that the combination of these two factors has made the subevents difficult to detect, rendering the lack of misordering detection an uninteresting artifact of the undectability of the subevents themselves. Therefore, in Experiment 3 we replicated Experiment 1 in one condition, and added a condition in which we asked participants to detect reversed subevent clips themselves and did not ask about detection of misorderings until the end of the experiment.

Experiment 3

In Experiment 3, we added a condition testing participants' ability to detect the critical out-of-order clip itself in each video. This "critical misordered clip" (see Figure 1) is the action that participants must see in order to detect the misordering. For example, in the action sequence illustrated in Figure 1 (correct: remove screw covering, pick up screwdriver, move screwdriver toward phone, insert screwdriver in screw; misordered: remove screw covering, move screwdriver toward phone, pickup screwdriver, insert screwdriver in screw) it is necessary that participants see the actor picking up the screwdriver from the table in order to determine that the action sequence was out of order. If participants do not detect this clip, it appears that perhaps an action is missing, but that the sequence appears in the proper order (they would just see the sequence: remove screw covering, move screwdriver toward phone, insert screwdriver in screw).

In addition to testing for detection of the critical misordered event, we tested participants' incidental detection of misordered action sequences via a brief posttest questionnaire, and again tested the effects of verbal interference both on detection of the critical clip and on detection of the misorderings.

Method

Participants. Thirty-four participants (25 female; *M* age = 19.03 years, *SD* = 0.99) were recruited from Vanderbilt University in exchange for course credit. Seventeen participants completed each condition (misordering detection and clip detection).

Stimuli and procedure. Participants viewed the same 12 videos shown in Experiment 1, and the videos were presented using the same equipment as Experiment 2. The misordering-detection condition was identical to that in Experiment 1. In the clip-detection condition, each video was preceded by the critical misordered clip. That is, the second of the two actions involved in the misordered act, as illustrated in Figure 1.

Each trial in the clip detection condition began with a 5,000-ms countdown, followed by a 1,000-ms black screen and then the critical misordered clip. After viewing the critical clip, the countdown and black screen were repeated, followed by the full video. Before the experiment began, participants were instructed to press a key on the keyboard as soon as they saw the critical misordered clip (described to them as the target clip) appear in the subsequent video. Reaction times (RTs) were measured from critical clip onset using timeline markers in Final Cut Pro. If multiple responses were made during one trial, the latest RT was selected for analysis. Participants in the clip detection condition were told only to detect the target clips and were not told that any of the videos contained an out-of-order action sequence. In both the clip-detection and misordering detection conditions, participants did the same interference task as in Experiment 1.

After completing the 12 trials, participants in the clip-detection completed the posttest questionnaire used in Experiment 2 to assess incidental misordering detection. As the questionnaire was not administered until the end of the experiment, participants were asked once about all 12 videos, rather than once for each individual video. Because of this, a participant only needed to detect one of the six misorderings across the 12 trials to be counted as having detected a misordering.

Results

Clip detection. Participants in the clip-detection condition were able to detect the vast majority of clips, even under conditions of interference. Participants detected 96% of target clips overall. In the interference condition, participants detected 94% of target clips, and in the no-interference condition they detected 99% of target clips ($p = .37$; McNemar test).

In addition, we found that participants sometimes anticipated clips, responding either before the actual event was visible, or responding very quickly after the onset of the event (100 ms or less after the onset of the clip). These early responses happened primarily on the misordered trials (occurring on 24.5% of misordered trials as opposed to 3.9% of nonmisordered trials), $t(16) = 4.440$, $p < .001$, and were usually followed by a repeat response (early responses with no subsequent response occurred on 8.8% of trials for the misordering trials, and on none of the no-misordering trials). Mean RT on hit trials (that did not elicit an early response) was 825 ms for the interference trials, and 612 ms for the no-interference trials, $t(16) = 2.564$, $p = .021$.

Misordering detection. The misordering-detection condition very closely replicated the findings in Experiment 1. Participants detected significantly fewer misorderings in the interference condition (25%) than in the no-interference condition (45%), $t(16) = 2.281$, $p = .037$. There was no difference in the prevalence of correct rejects between conditions (interference condition, 80% correct, no-interference condition, 86% correct, $t < 1$). This produced a d' score of .95 ($c = .603$) in the no-interference condition, and .167 (c = .758) in the interference condition. Overall accuracy was no greater than chance in the interference condition (52% accuracy) and it was significantly greater than chance in the no-interference condition (66% accuracy), $t(16) = 3.241$, p = .005. The difference in overall accuracy between conditions was a nonsignificant trend, although it should be pointed out that these are two-tailed tests, $t(16) = 1.875$, p = .079.

In addition to testing intentional detection of misorderings in this experiment, it was also possible to assess incidental misordering detection in the clip detection task because we asked participants whether they had seen any misorderings after they had completed all 12 trials, and if so, to describe them. Of the 17 participants in the clip-detection condition, three reported seeing one misordering each, although one of those three could not recall any details about what they had seen. The other two participants each reported having seen one misordering and were able to describe roughly when it occurred. Because there were six misorderings in each participant's trial sequence, this implies that three out of 102 misorderings were spontaneously detected.

Discussion

Experiment 3 demonstrates that *presence* of critical misordered clips was easily detected even when the fact that they were out of order was almost never detected. That said, there were a fair number of anticipatory responses. Most of these were subsequently corrected by an additional response once the actual clip appeared, but the phenomenon does constrain our hypotheses about prediction during event perception. It was telling that these responses were much more likely during reversed events. This makes sense because the critical clip (e.g., grabbing a screwdriver) was strongly suggested by the clip preceding it (holding the screwdriver) for the reversed sequences. Therefore, participants appear to have false-alarmed, either because they thought they missed the critical clip upon seeing action that should have followed it, or because they experienced some kind of deeper perceptual inference or filling-in and falsely perceived the action in the critical clip. In either case, this does suggest some form of active construction of the event sequence, although the details of the process are not clear.

One potential issue with Experiment 3 is that the critical clip detection task did not include nontarget videos that lacked the clip. Although participant responses were time-locked to the target clips, it remains possible that there would have been a high proportion of false alarms in the task, thus reducing the effective accuracy of clip detection. To further verify that individual clips were detectable in the absence of excessive false alarms, we ran an additional clip detection control experiment in which each of 20 participants searched for target clips in six videos that contained them and six videos that skipped them. In this experiment, the hit rate was 98.3% and the false alarm rate was 18% ($d' = 3.035$, $C = -.602$). This compares with hit rates of 53.3% and 45% for misordering detection in the no-interference conditions of Experiments 1 and 3, and false alarm rates of 6.7% and 14%, respectively ($d' = 1.58$, $C = .708$. and $d' = .955$, $C = -.615$, respectively). Because the false alarm rate for the critical clip detection control was higher than the comparable rates for the misordering detection we assessed the rates for individual videos and found that two of the videos had distinctively high false alarm rates of 80% and 70%, respectively. Inspection of these videos revealed shots very similar to the critical target shots that likely induced the very high false alarm rates. Removing these videos from the analysis results in a false alarm rate of 7% and a hit rate of 98% ($d' = 3.53$, $C = -.289$). Removing these videos from the calculation for the two misordering detection experiments results in hit rates of 61% and 56%, and false alarm rates of 7% and 19% for Experiments 1 and 3, respectively ($d' = 1.702$, $C = .569$ and $d' = 1.024$, C = .366). This analysis provides converging evidence against the possibility that difficulties in misordering detection occurred because the critical shots themselves were undetectable.

Finally, it is important to emphasize that this experiment clearly replicated the finding in Experiment 1 that interference lessened performance in reversal detection to chance levels. In addition, this experiment again demonstrated that spontaneous detection of reversals almost never occurs, even in cases where participants have clearly detected the reversed clip. This means that all participants reported a number of reversed clips (and sometimes even had to correct an anticipatory response to it) but nonetheless failed to register that it followed a clip that would have been out of sequence.

Experiment 4

As stated in the introduction, Claus and Kelter (2006) conclude that "Mental representation of the time course of a dynamic situation is a prerequisite for understanding" (p. 1042). Although Experiments 1–3 have demonstrated that detecting misorderings is more difficult than this hypothesis implies, it remains possible that between-cut blanks and rapid editing in our videos have made the events as a whole difficult to understand. If so, then a lack of understandability makes a failure to detect misorderings irrelevant to any hypothesis that default awareness of event sequence is necessary for event understanding. Therefore, in Experiment 4 we assessed the degree to which participants could understand and summarize the events portrayed in our 12 experimental videos, and whether relatively poor misordering detection in any given video might also be associated with poor event understanding.

In Experiment 4 we created new versions of each video that were slower, had fewer cuts and had no blanks between the few cuts that remained. We showed these new videos to one group of participants and the videos from Experiments 1–3 to another group of participants and tested the degree to which each group of participants was able to effectively describe the events in the videos.

Method

Participants. Thirty-two undergraduate students (24 female; M age = 18.69 years, SD = 1.03) were recruited from Vanderbilt University. Participants were compensated with course credit.

Stimuli and procedure. Reedited versions of the 12 original videos were created. The goal of these changes was to make the editing and speed of actions appear more typical, in order to determine whether the atypical editing and increased action speed decreased event understanding in previous experiments. First, the 67-ms black screen between each action was removed. Second, the videos were reedited to decrease the number of cuts. This step was necessary in order to reconstruct the videos, as the black screens were often interspersed across multiple actions shot from similar angles, and simply removing the black screens without further reediting would have resulted in highly unnatural jumps both across time and in the position of objects or actors. Finally, the sped up clips in the original videos were restored to 100% speed. These new videos contained an average of 2.59 shots ($SD = 2.02$), each with an average length of 14.16 s ($SD = 10.00$). This compares with an average of 11.25 shots per video and an average length of 9.58 s.

The participants were only shown videos containing the correct order of actions—no videos contained an out-of-order action sequence. Half of the participants watched the same versions of these 12 videos as had been shown in Experiment 1. The other half of the participants saw the new, reedited versions of these videos. Participants were instructed to write a summary of each video and were told to focus on the actions that made up the event. Summaries were written immediately following each video, and participants were given as much time as they desired to write their summaries. A naive rater matched participants' responses against a list of actions performed in each video (two to six actions per video), and this was compared with the first author's matching judgments producing very high reliability ($\kappa = .957$). The rater indicated whether the participant did or did not report the actions for each video, via a binary yes/no rating for each individual action. The number of yes ratings and the total number of possible yes ratings were used to calculate a percent accuracy score for each trial.

Results

Participants correctly recalled the vast majority of actions in both videos, but recalled slightly more actions for the new typically edited videos (91.20%, $SD = 6.20$) than for the old atypically edited videos (87.30%, $SD = 7.90$), $t(31) = 2.29$, $p = .03$. Although participants generally performed better in the typically edited video condition, average performance was better in the atypically edited video condition for four of the 12 videos.

Because there was a small difference in comprehensibility between the conditions, we tested whether relatively poor comprehensibility in a given video would be associated with poor reversal detection. Misordering detection performance for each individual video was compiled from past experiments. These experiments include no-interference misordering detection results from Experiments 1 and 3, along with three additional experiments not included in this article. Two of these experiments tested manipulations of the length of misordered clips, and of clips preceding and following misordered clips. These manipulations did not produce consistent results. The third was a basic misordering detection experiment that produced results very similar to those reported in Experiment 1. We considered only trials (a) with an out-of-order action, (b) using the original video stimuli with no extra tasks (mimicking the no-interference trials of Experiment 1), and (c) using the same population and presentation methods as the experiments described here. Using these criteria, we were left with 309 total responses from 100 participants (M age = 19.47, 72 female), and 24 to 29 reversal detection responses per video. Across the 12 videos, there was a nonsignificant negative correlation between percent correct responses on the misordering detection task and percent correct responses on the event recall task for trials containing the original atypically edited videos ($r = -.36$, $p = .25$). This provides further evidence that the original misordering detection results were not related to any difficulty in comprehending the videos.

Discussion

Experiment 4 demonstrates that the events portrayed in the videos used in Experiments 1–3 were clearly understandable. Not only did viewers correctly describe most of the events in the video, but they were very nearly as accurate for the speeded videos as for the unspeeded videos. In addition, there was no correlation across videos between description accuracy and reversal detection. Therefore, these results demonstrate that incomprehensibility did not produce the reversal detection failures we have observed.

Experiment 5

These experiments demonstrate that detecting event reversals is difficult, and that it requires limited capacity cognitive resources. In Experiment 5 we provide initial evidence demonstrating how these resources are used. On one view, perhaps additional resources are used to generate more effective predictions in which relatively deeper elaboration on one event can generate more detailed predictions for the next event. Although we believe that this likely occurs, our goal in these initial studies is to establish the plausibility of the alternative that events are understood using both predictive processes and more postdictive assembly processes in which current events are compared with previous events. These operations may require resources during a time following a given event, and therefore one of the primary determinants of poor reversal detection may be the resource load necessitated by understanding the events following the reversal. Therefore, in Experiment 5 we tested the degree to which reversal detection would be improved by eliminating the events that occur subsequent to the reversal.

Method

Participants. Twenty-four undergraduate students (19 female; M age = 21.5 years, $SD = 6.53$) were recruited from Vanderbilt University and the surrounding community. Participants were compensated with course credit and payment.

Stimuli and procedure. All videos were presented as in Experiment 2 with the exception that half of the videos each participant viewed ended immediately after the critical out-of-order clip was displayed. Due to the nature of the videos, this change required normal order videos in this condition to contain one fewer clip than videos containing a misordered action, as the critical clip occurred one action earlier in these videos. Participants were instructed to detect misorderings and again responded after each clip.

Results

Participants were significantly more accurate (84.1% correct) when the video ended immediately after the reversal than when the video continued as in the previous experiments, 75.0% correct; $t(22) = 2.703$, $p = .013$. For target-present videos, participants detected significantly more reversals in the early ending videos (73.9%) than in the full videos (55.1%), $t(22) = 2.614$, $p = .016$, and there was no difference in correct rejections (94.2% correct rejections for the short videos, and 95.6% for the full videos, $t < 1$). This produced d' scores of 2.21 ($c = .466$) for the early ending videos, and 1.83 ($c = .788$) for the full videos.

Discussion

Experiment 5 demonstrates that events following the reversal play an important role in making misorderings difficult to detect. This implies that at least some of the resource-intensive processing necessary to detect the reversals occurs after the reversal. Of course, it might be possible to argue that some form of masking has been eliminated in this experiment, but the operations necessary to detect reversals make this comparison awkward at best. Traditionally, a backward mask needs to occur within 50 ms of the masked stimulus to be effective, and some rapid feed-forward process must be hypothesized that can limit awareness and basic processing for the immediately preceding stimulus (for a review, see Breitmeyer & Öğmen, 2000), and related phenomena such as the attentional blink lessen detection of targets if a subsequent target occurs within about 500 ms (for a review, see Shapiro & Luck, 1999). In contrast, reversal detection requires some level of encoding for an initial subevent lasting 1–2 s, followed by some form of comparison with a second subevent which itself lasts 566–1,066 ms in the present experiments. Then, a 67-ms blank screen follows, and for the continuing events additional actions follow that. These parameters are more consistent with forms of conceptual masking that leave basic perception of stimuli intact but interfere subsequent target recognition (Loftus & Ginn, 1984; Potter, 1976). More conceptual forms of masking can be effective for up to 500 ms (Potter & Levy, 1969), or possibly somewhat more (for a review, see Potter, 1999) which approaches the duration of our briefest misordered shots. These limits are particularly interesting here because they implicate an apparently default identification/elaboration process that appears to compete for limited-capacity resources with the comparison processes that would seem necessary for sequence perception.

Based on these considerations, it seems likely that improved performance in the short-video conditions occurred because viewers were able to use the moments after the misordering to process the relevant events without interference from the need to continue the reversal search, or at least to perceive the subsequent events. Because the absence of postmisordering event information increased detection, this finding strongly implies that correctly perceiving the sequence of events requires reconstructing the recent past, as new information must be checked against previous events. Evidence from all of the experiments in this article suggest that this is a resource intensive process, and Experiment 5 suggests that this process might require just as much looking backward as forward.

General Discussion

The experiments in this article have demonstrated that participants are frequently unaware of misorderings in action sequences. Experiments 1 and 3 demonstrate and replicate significant reductions in performance caused by a secondary verbal task, and Experiments 2 and 3 demonstrate and replicate near-zero levels of incidental detection of misorderings. However, even in situations where participants were unaware of misorderings, they were almost always able to detect the critical misordered event itself, both with and without an interfering task. Experiment 4 verified that the events were easily comprehensible, and that variation in comprehensibility was not associated with levels of misordering detection. Finally, Experiment 5 demonstrated that misordering detection accuracy increased when the sequence ended immediately after the critical event.

Several issues are important in contextualizing these results, but perhaps the most central is to consider the degree to which some more sensitive measure might reveal evidence that misorderings have affected event perception. For example, it is possible that events subsequent to a misordering (or the misordered events themselves) would be less primed than the same events following a correct sequence. Another possible effect of misordering would be to subtly disrupt a higher-level form of organization which might allow easier access to parts of events just as disrupting face configuration can allow easier access to face parts (Young, Hellawell, & Hay, 1987). A similar approach might be to test the degree to which some measure of performance suffers for reversed events. Given the close relationship between sequence prediction and segmentation hypothesized by event segmentation theory, it would be particularly interesting to test the degree to which viewers can successfully recognize, learn, or segment reversed sequences (especially after repeated presentations).

A segmentation experiment would be particularly interesting because misordering-induced prediction errors are assumed to induce segments. We have recently completed a research demonstrating that near-threshold spatial disruptions can induce increased event segmentation (Baker & Levin, 2015), and it is possible that misorderings would have a similar effect even if they go undetected. Alternatively, there could be no impact of misorderings on segmentation. Although this would be a difficult-to-interpret negative finding, it would reinforce the hypothesis that event sequence perception is not a necessary component event understanding.

However, even if misorderings were to induce segmentation, it is instructive to consider the nature of a system in which misordering-produced error signals can impact conscious detection of an event segment, while the misorderings can themselves escape notice. This would mean that the segmentation-induced analysis of actions, goals, and context-embedded reward processing does not access the sequence information that has created a nominal violation of some event schema, even in the moments after it has been present. Consider how prediction would work in the example illustrated in Figure 1. A viewer sees someone with a screwdriver in hand and would presumably predict that the next step would be to use it on a screw. Instead, a previous step is shown - grabbing the screwdriver. In a truly predictive framework, effective event understanding would require that some representation of “using a screwdriver” be activated and available in working memory before the subsequent scene appears. When the grabbing scene is shown, a comparison between the predicted event stored in working memory and the on-screen event produces an error signal which induces an event representation that brings to awareness something important about the nature of the event, and presumably induces increased encoding (this can be seen as the perceptual gating process in event segmentation theory; Zacks et al., 2007), all without affording awareness of the unusual sequence that led to the signal.

This would seem to be counterproductive because the unusual sequence information is precisely the kind of information one would think would be necessary to really understand this new event. Perhaps the actor has reached for new screwdriver because the one he was holding previously was not the correct size - this is a possible interpretation and it would seem that a system of constant prediction and comparison would be ideally attuned to detect it, or at least to bring it to awareness for further deep analysis. As such, our findings place important limitations on the effects of analyses that follow from prediction errors even if they are generated by default. They apparently cannot bring to awareness possible alternative event sequences, or even a more basic comparison between a working memory representation of an initial event and a currently visible event, all of which constitutes evidence against the hypothesis that sequence perception necessarily lies at the heart of event perception.

When considering the utility of either an event-segmentation experiment or a priming experiment, it is important to point out that no one measure can claim to be an exhaustive assessment of the impact a stimulus has had on a viewer, and that implicit measures clearly have interpretive issues of their own (for a review, see De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). However, this does make clear the need to consider how failures to become aware of misorderings, both in incidental tasks and intentional detection tasks, constrain explanations of event perception and ongoing visual awareness more generally.

At the broadest level, failures to detect reversals will have the most impact on our understanding of similarly conscious components of event perception. Given the importance that event segmentation theory places on goals as a basic feature organizing event models, research exploring how early developing theory of mind skills helps children and adults reason about goal-drive actions would seem particularly relevant when considering the possibility that much of event perception requires nondefault processing. A clear example of this need can be seen in situations described by Leslie, Freidman, and German (2004), who argue that false belief understanding requires an inhibitory process that guides attention away from an object that is the target of a true belief to the correct target of a false belief. So, in a classic false belief scenario (e.g., Baron-Cohen et al., 1985), a viewer sees another person witness an object being hidden in one of two locations. Then, the witness leaves the room and the object is moved to the other location. The now-ignorant witness returns to search for the object, and the viewer must predict where the witness will search.

From a visual event sequence point of view, one could easily imagine a slight variant of the false belief scenario in which the viewer sees the witness watch a thing being hidden, go out of the room, return, reach into the wrong hiding spot, look inside, look at the correct hiding location, reach into the correct hiding spot, and withdraw the object (similar events have been used to demonstrate simple forms of false belief understanding in infants; see, e.g., Kovács, Téglás, & Endress, 2010). Understanding this sequence requires a series of cognitions that includes representations of the witness's goals, endogenous attentional shifts to important objects, and, according to Leslie et al. (2004), inhibition of the correct hiding spot so that the viewer can generate a prediction (or at least an explanation) for witness's reach to the wrong hiding spot.

So, if this is the cognitive framework necessary to understand this well-structured event, then it is worth considering what could happen during this sequence to defy the viewer's cognitive model and require the kind of new-event updating hypothesized in event segmentation theory. For example, upon returning to the room, the witness could look into the box and then reach into it - this might imply that the witness returned with the suspicion that the object had been moved. Because this kind of sequence variation could be important, it requires an already complex situation be interpreted with a still deeper analysis and even simulation of multiple possible scenarios. In a case such as this, it seems unlikely that a simple automatic analysis of the relationship between events could suffice to differentiate an unexceptional continuation of an event from a novel change in goals. Accordingly, it would make sense that this sequence not be encoded by default, but rather is left to higher-level systems that engage an explicit analysis of the sequence only if the derived meaning of the events and the perceiver's goals demand it. Consistent with this, research reviewed above demonstrates that adult participants often fail to engage in the cognitions necessary to fully track beliefs (Barr & Keysar, 2005).

What this discussion suggests is that an analysis of event sequences that that can effectively test online predictions will be difficult to do "on the cheap" using perceptual routines or modules that lack access to the full range of knowledge necessary to interpret an event. This is not to suggest that viewers cannot do this ---our data clearly demonstrate that when viewers devote resources to the problem they can detect a reasonable proportion of reversals, and they rarely false alarm to other event-to-event transitions that occur in an expectation-consistent order. However, our data clearly demonstrate that this requires considerable effort, and Experiment 5 demonstrates that ongoing events subsequent to a reversal interfere with this effort. Accordingly, the benefit of detecting some proportion of actions that signal new events may be outweighed if it comes at the cost of alerts that are ineffective without capacity-intensive processing. Thus, it may make sense to leave many elements of sequences to remain in the outside world of events just as it is plausible that the visual world can serve as an "outside memory" for visual properties (O'Regan, 1992). After all, many actions constrain each other into sequences (e.g., you cannot use something before grabbing it) and so the difficult problem of monitoring them might profitably be limited to the minority of circumstances where they are likely to be important and therefore will generate a payoff for the application of limited-capacity resources.

However, if we are to argue that event perception sometimes involves ongoing prediction and comparison, but sometimes does not, it is important to specify when each of these approaches might be used. Accordingly, these findings could be developed to enhance theories of event perception that depend on the awareness of prediction errors (Zacks et al., 2007), and theories that assume that sequence is an inherent dimension of event perception (Claus & Kelter, 2006; Raisig et al., 2010). A number of interesting recent theories of event perception and executive functioning may provide a basis for this development. For example, Sirigu et al. (1996) argue that text comprehension can either occur in a sequential/predictive mode in which interevent relationships are preferentially processed, or a more hierarchical semantic mode in which the meaning of individual actions are processed. This hypothesis is based on interesting data suggesting a dissociation whereby patients with frontal lobe damage have difficulty with narrative sequencing tasks, but are more successful with more semantic tasks. With some modifications, this distinction might apply to visual event perception if one assumes that in some cases viewers focus their attention in interevent relationships, and would therefore be more likely to notice misorderings, whereas in other cases they are more focused on identifying events, a task for which sequence might be of secondary importance. In a similar vein, Braver (2012) argues that different kinds of executive functioning can be effectively characterized by a distinction between proactive and reactive processing. In the proactive mode, individuals organize their cognitions around goal maintenance and prospective inhibition, and it is easy to imagine how this might include representation and comparison of current and upcoming events. In contrast, the reactive mode reflects more of a "just in time" form of cognitive control characterized less by goal maintenance than by identification and execution of goal-relevant actions only when they are clearly needed. On this view, it seems possible that many forms of event perception, especially at the level of simple actions, could take advantage of something like a reactive mode which allows efficient event identification without the need for extensive resource-occupying continuous maintenance of representations of goals or other information about the current state of an ongoing event. Accordingly, the findings reported here can improve our understanding of event perception by making clear the need for efficiency in tailoring often-difficult representation and comparison processes to situations where they are useful.

References

Aginsky, V., Harris, C., Rensink, R., & Beusmans, J. (1997). Two strategies for learning a route in a driving simulator. *Journal of Environmental Psychology, 17,* 317–331. http://dx.doi.org/10.1006/jevp.1997.0070

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841–844. http://dx.doi.org/10.1111/j.1467-9280.2006.01791.x

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. *Cognition, 136,* 14–29. http://dx.doi.org/10.1016/j.cognition.2014.11.030

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106,* 1382–1407. http://dx.doi.org/10.1016/j.cognition.2007.07.005

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7,* 66–80. http://dx.doi.org/10.1162/jocn.1995.7.1.66

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21,* 37–46. http://dx.doi.org/10.1016/0010-0277(85)90022-8

Barr, D., & Keysar, B. (2005). Mindreading in an exotic case: The normal adult human. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 271–283). New York, NY: Guilford Press.

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

Breitmeyer, B. G., & Öğmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62,* 1572–1595. http://dx.doi.org/10.3758/BF03212157

Claus, B., & Kelter, S. (2006). Comprehending narratives containing flashbacks: Evidence for temporally organized representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1031–1044. http://dx.doi.org/10.1037/0278-7393.32.5.1031

Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology, 8,* 196–209. http://dx.doi.org/10.5709/acp-0115-z

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135,* 347–368. http://dx.doi.org/10.1037/a0014211

Eagleman, D. M., & Sejnowski, T. J. (2000, March 17). Motion integration and postdiction in visual awareness. *Science, 287,* 2036–2038. http://dx.doi.org/10.1126/science.287.5460.2036

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 126–132. http://dx.doi.org/10.1037/0278-7393.10.1.126

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

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7,* 43–64. http://dx.doi.org/10.1080/135062800394676

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12,* 822–851. http://dx.doi.org/10.3758/BF03196775

Keen, R. (2003). Representations of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science, 12,* 79–83. http://dx.doi.org/10.1111/1467-8721.01234

Kerzel, D., & Gegenfurtner, K. R. (2003). Neuronal processing delays are compensated in the sensorimotor branch of the visual system. *Current Biology, 13,* 1975–1978. http://dx.doi.org/10.1016/j.cub.2003.10.054

Koenderink, J. (2012). *Chrono-geometry*. Retrieved from www.gestaltrevision.be/pdfs/koenderink/ChronoGeometry.pdf

Kolers, P. A., & von Grünau, M. (1976). Shape and color in apparent motion. *Vision Research, 16,* 329–335. http://dx.doi.org/10.1016/0042-6989(76)90192-9

Kovács, A. M., Téglás, E., & Endress, A. D. (2010, December 24). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330,* 1830–1834. http://dx.doi.org/10.1126/science.1190792

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences, 8,* 528–533. http://dx.doi.org/10.1016/j.tics.2004.10.001

Loftus, G. R., & Ginn, M. (1984). Perceptual and conceptual masking of pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 435–441. http://dx.doi.org/10.1037/0278-7393.10.3.435

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes, 22,* 199–224. http://dx.doi.org/10.1080/01638539609544973

McBeath, M. K., Shaffer, D. M., & Kaiser, M. K. (1995, April 28). How baseball outfielders determine where to run to catch fly balls. *Science, 268,* 569–573. http://dx.doi.org/10.1126/science.7725104

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

Nijhawan, R. (1994, July 28). Motion extrapolation in catching. *Nature, 370,* 256–257. http://dx.doi.org/10.1038/370256b0

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Review Canadienne de Psychologie, 46,* 461–488. http://dx.doi.org/10.1037/h0084327

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. http://dx.doi.org/10.1037/0278-7393.2.5.509

Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13–46). Cambridge, MA: MIT Press.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10–15. http://dx.doi.org/10.1037/h0027470

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4,* 515–526. http://dx.doi.org/10.1017/S0140525X00076512

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2007). Investigating dimensional organization in scripts using the pupillary response. *Psychophysiology, 44,* 864–873. http://dx.doi.org/10.1111/j.1469-8986.2007.00592.x

Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology, 76,* 1–8. http://dx.doi.org/10.1016/j.ijpsycho.2010.01.006

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7,* 345–376. http://dx.doi.org/10.1080/135062800394847

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8,* 368–373. http://dx.doi.org/10.1111/j.1467-9280.1997.tb00427.x

Ruby, P., Sirigu, A., & Decety, J. (2002). Distinct areas in parietal cortex involved in long-term and short-term action planning: A PET investigation. *Cortex, 38,* 321–339. http://dx.doi.org/10.1016/S0010-9452(08)70663-4

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18,* 37–55. http://dx.doi.org/10.1002/acp.940

Schwan, S., & Garsoffky, B. (2008). The role of segmentation in perception and understanding of events. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 391–414). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195188370.003.0017

Shapiro, K. L., & Luck, S. J. (1999). The attentional blink: A front-end mechanism for fleeting memories. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 95–118). Cambridge, MA: MIT Press.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7,* 1–15. http://dx.doi.org/10.1080/135062800394658

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5,* 644–649. http://dx.doi.org/10.3758/BF03208840

Sinico, M. (1999). Benussi and the history of temporal displacement. *Axiomathes, 10,* 75–93. http://dx.doi.org/10.1007/BF02681817

Sirigu, A., Zalla, T., Pillon, B., Grafman, J., Agid, Y., & Dubois, B. (1996). Encoding of sequence and boundaries of scripts following prefrontal lesions. *Cortex, 32,* 297–310. http://dx.doi.org/10.1016/S0010-9452(96)80052-9

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

Thornton, I. M., & Hayes, A. M. (2004). Anticipating action in complex scenes. *Visual Cognition, 11,* 341–370. http://dx.doi.org/10.1080/13506280344000374

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3,* 86–94. http://dx.doi.org/10.1167/3.1.9

Vinson, N. G., & Reed, C. L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition, 9,* 41–65. http://dx.doi.org/10.1080/13506280143000313

von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Tracking and reaching for moving objects. *Cognition, 67,* 255–285. http://dx.doi.org/10.1016/S0010-0277(98)00029-8

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception, 16,* 747–759. http://dx.doi.org/10.1068/p160747

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., . . . Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4,* 651–655. http://dx.doi.org/10.1038/88486

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,* 4057–4066. http://dx.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,* 273–293. http://dx.doi.org/10.1037/0033-2909.133.2.273

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

Received July 4, 2014
Revision received April 24, 2015
Accepted April 27, 2015 ■