The Influence of Sequence Reversal on Event Perception

Yining Ding

Research Mentor: Dr. Daniel Levin

Vanderbilt University

Abstract

Many theories in event perception suggest that the information about the temporal organization of events plays an important role in facilitating the comprehension of event content. Although a previous study conducted by Hymel et al. (2016) showed that most people were not aware of the presence of misordered events while viewing live-action videos of everyday activities, the current study aimed to use more sensitive measures, such as event memory and event segmentation, to reveal the impact created by sequence reversal on the perception and representation of events. In the experiments reported here, we discovered that viewers did not encode more visual details when the misordered event happened. The presence of reversals impaired viewers' ability to remember the location of the current event in the general event sequence, but this effect disappeared when viewers engaged in an event segmentation task and detected reversals incidentally. In addition, the existence of event misorderings did not increase the number of event boundaries experienced by the viewers. These results reinforce the idea that viewers do not engage in moment-to-moment examination of event sequence as a default process. We argue that even though there is evidence that the reversal exerts an influence on viewers' lower-level processing, reversals are rarely brought into conscious awareness and minimally impact viewers' mental representation of events, especially when there is no task-specific demand to focus on event sequence.

Keywords: event perception, dynamic visual cognition, sequence perception, event memory, event segmentation

The Influence of Sequence Reversal on Event Perception

The reality that we perceive and interact with consists of numerous events. For example, when being asked to describe our morning routine, we may refer to the process as a series of discrete steps that are bounded by sequential orders: Getting up, making the bed, going to the bathroom, etc. Each step can be considered as a separate event that we isolate from the continuous flow of consciousness, and each event is defined in psychological terms as “a segment of time at a given location that is conceived by an observer to have a beginning and an end” (Zacks & Tversky, 2001, p. 17). In order to effectively construct the reality that we perceive, it is natural to believe that we need to represent not only each individual event, but also the temporal order that governs the structure of the events.

A wide range of theoretical findings suggest that representing event sequence correctly is crucial for perceiving and understanding events. According to the Event Segmentation Theory (EST) proposed by Zacks et al. (2007), a correct representation of event sequence is the key to an error-driven prediction mechanism that helps us segment our experiences into discrete events: While viewing an event, people store multi-sensory information of “what is happening now” using working memory representations called the event model. The creation of each event model is also influenced by previously acquired knowledge about the temporal structure of familiar activities, which is called event schemata. Based on the current event model, people continuously generate predictions about the next state of the ongoing event. When these predictions conflict with the new incoming information, an increase in prediction errors causes the current event model to update itself to reflect new features in the environment, which results in our subjective experience of event boundaries. Supported by behavioral, neurophysiological, and computational evidence (Zacks et al., 2011; Reynolds et al., 2007), the predictive mechanism in this model not only confirms the importance of sequence information in event segmentation, but also implies that viewers should be consciously aware of each mismatch-induced prediction error that occurs, since the presence of these errors triggers the reconstruction of their current event model in working memory. Other empirical evidence also suggests that representing sequence and detecting sequence violations might be a default process in event perception: Using narrative comprehension tasks, Claus and Kelter (2006) discovered that despite the lack of explicit temporal structures in the text, readers were still able to generate temporally organized representations of events in the text to facilitate their comprehension. They explicitly argued that "mental representation of time course of a dynamic situation is a prerequisite for understanding" (p. 1042). Furthermore, by showing that a violation in the temporal order of visual sequences elicited a greater pupillary response in people than normally ordered sequences, Raisig et al. (2010) maintained that detecting sequence misordering is crucial in the process of understanding an event.

Many previous studies have explored the possibility that sequence processing can result in higher-level learning processes that exist outside of awareness. For example, when experiencing simple event sequences, the ability to extract information regarding statistical contingencies and utilize this information in other cognitive processes is referred to as implicit statistical learning. One of the earliest studies on this topic was done by Saffran et al. (1996), who used pseudospeech streams to examine 8-month-old infants' ability to extract key information from speech inputs. Saffran et al. discovered that infants were able to apply the skill of discriminating familiar syllable pairs to segment continuous speech streams and mark word boundaries, in an effort to facilitate language acquisition (Aslin et al., 1998). Other studies demonstrated that the skill of auditory statistical learning can also be useful in non-linguistic perception processes, such as listening and understanding musical tones (Creel et al., 2004).

The implications of statistical learning have also been generalized to the visual domain. Fiser and Aslin (2001) demonstrated that people can extract spatial correlations from scenes that contain multiple elements using only statistical information (i.e. which configuration occurred more frequently). In the visual temporal domain, Fiser and Aslin (2002b) also found that people can use statistical learning to extract the temporal relationship among shapes that were presented sequentially (Fiser & Aslin, 2002a). Although the participants received no specific instruction while viewing a computer animation of shape transformations, they were able to implicitly acquire the knowledge that certain shapes always occurred as sequences of triplets and could identify these triplets as more familiar in a later familiarity judgment task. Despite using laboratory stimuli like streams of simple geometric shapes, this study indicated the possibility that while viewing an event, people can unconsciously extract information about temporal organization and utilize this information to facilitate their understanding of the event.

In more naturalistic scenarios of event perception, many empirical findings have stressed how representing event structure helps with the processing of event contents. According to the event segmentation theory, the spatiotemporal location of each event serves as the basis of the creation of an event model (Radvansky & Zacks, 2011). In particular, event sequence may serve the role of “anchors for memory” (Richmond et al., 2017, p. 114). When movies were edited into summaries that omitted critical event boundaries, people’s recall of content in the movie became worse than when they saw movies summaries that retain these critical breakpoints (Schwan & Garsoffky, 2004). In addition, long-term associative memory for within-event information has been found to be better than information across event boundaries, which has been proved in both narrative contexts (Ezzyat and Davachi, 2011) and dynamic visual contexts (Kurby and Zacks, 2021). Consistent with these findings, many studies have suggested a positive correlation between event segmentation ability (i.e. the degree to which the locations of event boundaries decided by people correspond with the result of a larger group) and episodic memory performance (Kurby & Zacks, 2011; Sargent et al., 2013; Zacks et al. 2006).

Therefore, it was surprising when a direct assessment on people's awareness of event sequence using naturalistic stimuli revealed that viewers might not be so sensitive to violations of event sequence. A study by Hymel et al. (2016) investigated the degree to which audiences were aware of event sequence reversals when watching videos that depict familiar everyday events. In the study, each participant watched several videos containing events that did not follow the conventional sequence. For example, in the normal version of a video about an actor using a screwdriver on a phone, the actor would be seen picking up the tool from the table, using it on the phone, and then opening the phone successfully. But in the misordered version, two events that were supposed to be sequentially bounded switched position: The actor would be seen picking up the screwdriver from the table after they had used it on the phone. Hymel et al. discovered that misordering detection was difficult (only 53% misorderings were detected) when the participants were explicitly told to detect misordered events, and the task was almost impossible when there was no explicit instruction or when an interference task demanded attention at the same time. However, in a task that measured their understanding of the video, participants were able to write accurate video summaries and recall most events that happened, despite the fact that they were not aware of the sequence reversals.

The result of the study by Hymel et al. (2016) provided evidence that the representation of event sequence may not be generated by default, since the task of identifying temporal violations required limited-capacity attention, and the viewers did not seem to identify sequence violation spontaneously. However, it is also possible that the measure of awareness used in the Hymel study underestimated the importance of event sequence representations. If unusual event sequence information works implicitly to shape viewers' perception of events, even though viewers do not seem to be aware of the sequence violations, these disruptions in event sequence may still affect the way events are represented in their mind.

In this study, we built on the work of Hymal et al. (2016) and used more sensitive measures to examine how the presence of sequence reversal might influence the way people perceive and understand visual events without causing conscious experience of the reversal itself. One of the sensitive measures we used was the memory of visual details, because many previous studies have argued that working memory or short-term memory for visual events might work as an unconscious processing mechanism (Soto et al., 2011; Brogaard, 2011). The other measure was event segmentation, which has long been established as an automatic process alongside event perception (Radvansky & Zacks, 2011).

Based on the predictive mechanism in the Event Segmentation Theory, if viewers do constantly compare the temporal structure they experience at the moment with the prediction they generate from the current event model, sequence reversals should be the type of information that induces more prediction errors in their online processing, which will subsequently lead to an update of the current event model and a subjective experience of event boundary. If, as we hypothesized, unusual sequence information is able to drive this process without eliciting conscious awareness of the reversal itself, we might observe the following behavioral signatures: First, since the process of updating an event model requires a transient increase of sensory input to the current event representation (Zacks et al., 2007), viewers' ability of encoding detail information perceived during the reversal will be enhanced. Second, since the ultimate goal of updating an event model is to form a more accurate representation of what happens in the event (Radvansky & Zacks, 2011), a reversal-induced event model updating may allow the viewer to remember the event sequence more accurately. Third, as was stated by the core assumption of the Event Segmentation Theory, an increase in prediction errors caused by the presence of reversals may increase the number of event boundaries experienced by the viewer. A similar effect was observed by Baker and Levin (2015) on the impact of discontinuous spatial relationships, suggesting that the presence of relational triggers can lead to the perception of multiple event boundaries. We examined the above three hypotheses in the following two experiments.

Experiment 1

In Experiment 1, we used a newly-created set of video stimuli to replicate the low detection rate of sequence reversals in the study by Hymel et al. (2016) and employed two memory tasks to examine the first two hypotheses above: If the presence of sequence reversals can trigger event model updating and the formation of a more accurate new event model, one would expect to observe that viewers' encoding of visual details during the reversal increases, and that they perform better in remembering the overall event sequence. In addition, since this process would not require a conscious awareness of the reversals, there would be no correlation between reversal detection ability and these two types of memory performances. This hypothesis was supported by previous research suggesting that different visual features of the same entity were encoded independently during perception (Fougnie & Alvarez, 2011). To ensure that the memory questions used in the study were not excessively difficult and that our viewers actually used the information from the videos to answer them, we included a control group in which participants answered the questions without watching the videos, thereby establishing a chance-level performance baseline for both visual detail questions and event sequence questions.

Method

Participants

87 Vanderbilt University students participated in the study in exchange for course credit. Nine of them were excluded for completing only part of the survey (the progress tracked in Qualtrics showed less than 100% completion), leaving 78 records being used in the analysis (48 female, average age = 19.5 years, SD = 1.2 years). Among these 78 records, 58 participants completed the main tasks in the experimental group, and 20 participants were in the no-video control group. The experiment was approved by Vanderbilt University's institutional review board, and all participants completed informed consent forms prior to the experiment.

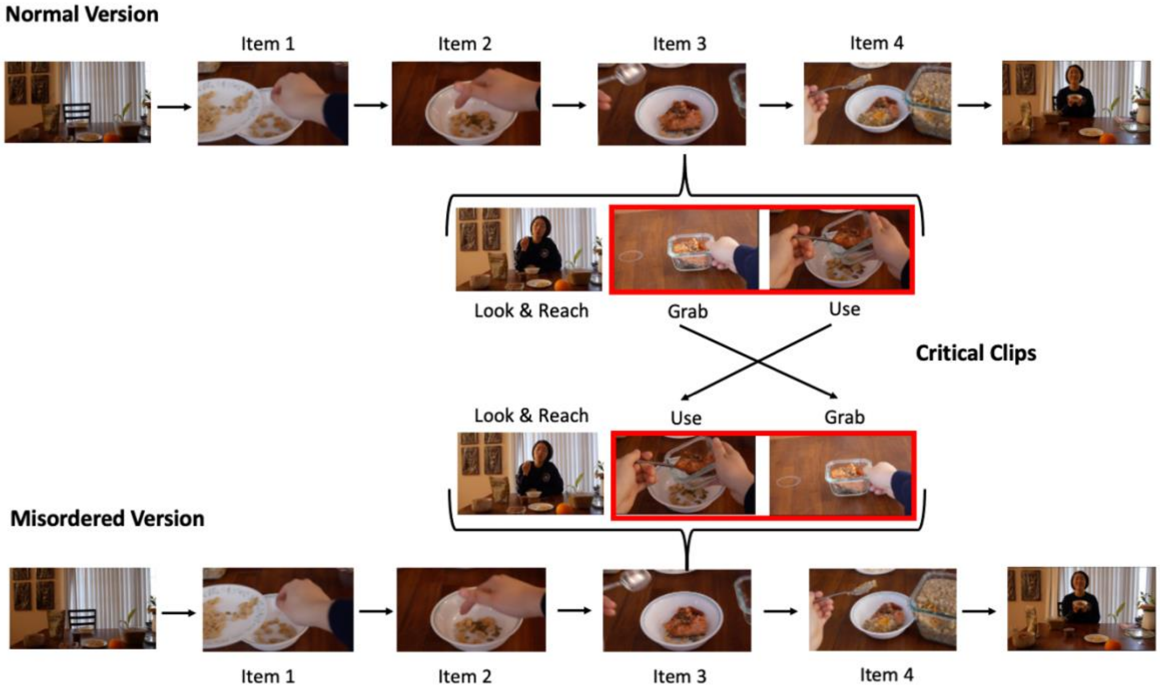
Stimuli

The stimuli used in the experiment were self-created live-action movies similar to those in the study by Hymel et al. (2016), but they did not contain blanks between shots. These movies featured one actress performing familiar daily activities, including someone building blocks with Lego (32 s), making a breakfast bowl (20 s), mixing ingredients for a porridge (26 s), assembling a violin (32 s), preparing to go out (14 s), and making a fruit salad (10 s). Each movie was composed of a series of medium shots and close-ups showing the actions of the actress and the objects she used. The average length of the movies is 22.5s.

To manipulate whether each video seen by the subject contained misordered events, we edited each video into a normal version and a reversed version using iMovie. In the normal version of videos used in this experiment, the actress was seen completing a goal (e.g. making a breakfast bowl) by manipulating a series of different items (e.g. adding peanuts, pumpkin seeds, salmon, and oatmeal to the bowl). For each item she manipulated, the actress was seen first (1) looking at and reaching out to it, then (2) grabbing it, and then (3) using it (e.g. put it into the bowl). But in the reversed version after editing, the action sequence of her manipulating one of the items was disrupted: For the action performed on the critical misordered item (e.g. salmon), she was seen first (1) looking at and reaching out to it, then (3) using it, and then (2) grabbing it, and the action sequence of all the other items remained normal (See Figure 1). The two clips of "Grab" and "Use" of the critical misordered item were referred to collectively as "critical clips". Across all the six movies used in the experiment, the mean duration of the critical clips is 1.6s (SD = 0.44s).

Figure 1

Examples of Normal and Misordered Versions of Videos



Note. Images are screenshots from one of the videos used in the experiment.

Design

Overall, this experiment used a within-subject design. The independent variable was the existence of sequence misordering in the video. For participants assigned to the experimental group, half of the videos they saw were in misordered version, and half were in normal version.

In the experiment, we measured viewers' memory for two types of information: The first type is visual details in the critical misordered clips (i.e. the "grab" and "use" clips). According to Hymel et al. (2016), although our stimuli in misordered version incorporated two event types that might confuse participants' prediction of event sequence, namely one ellipsis (e.g. from "look & reach" to "use, without "grab" in between) and a return to an action that should be done earlier (e.g. first "use", and then "grab"), only the later type was more anomalous to viewers and would directly cause disruption in their perception of event sequence. Therefore, it was natural to assume that if the existence of misordering influenced people's memory of event details, its impact should be limited to the scene when they saw a return to a previous action, namely the critical clips. As an example, one sample statement that was judged by the participant in this section was "When she grabbed the salmon, a plate was visible next to it." The other measurement was the memory of the sequential order of events in the whole video, specifically the serial position of the critical event in regard to other events that happened in the video. For example, a sample statement was "The orange salmon was the second ingredient she put into her breakfast bowl." In order to ensure that the difficulty of these statements regarding event details was appropriate, we also established a control group, in which participants did not watch the video, but were nonetheless tested on all the questions regarding the videos. Their performance was compared against participants in the experimental group to ensure the later group used the visual information from the videos to answer these questions.

We also measured the viewers’ ability to detect reversals, which was defined in the study as participants’ awareness of sequence misorderings in the event. After watching each video, participants were asked to report whether they saw anything out of sequence and identify what the item being manipulated was when the misordering happened. Compared to the video stimuli used in the study by Hymel et al. (2016), the videos used in the current study did not contain blanks in between shots. Therefore, we would like to see if the reversal detection ability remained low without adding extra cognitive load for the subjects to encode sequence information.

Procedure

Participants accessed the experiment online using Qualtrics. Before they began the experiment, participants were instructed to provide some demographic information, including their age and gender. Then they read an instruction containing the definition of a “reversal” in the video and their task during the experiment: “For some of the videos you will watch, all of the actions will be in the correct order, but in other videos some action will occur out of order (E.g., ‘Grabbing a can from the refrigerator’ happens before ‘opening the refrigerator’). Your job is to pay close attention to detect the misorderings and to attend to details in the video.” Then they were tested on whether they remembered the key information by selecting the statement that did not occur in the instruction.

Each participant was randomly assigned to an experimental group or a control group. Participants assigned to the experimental group watched three normal videos and three videos containing misorderings. The order of the six videos and their accompanying questions was randomized for each participant. Before each video started, there was a countdown for five seconds. All videos were played at a resolution of 640 * 360 on the participants’ own digital device. After watching each video, the participants were instructed to move on to reversal detection questions and true-or-false judgement questions. For the reversal detection questions, the first question asked whether they detected anything out of order in the video. If the answer was yes, they were required to specify which object was being handled when the reversal happened; and if the answer was no, they typed “N/A”. On the next page, there were seven true-or-false judgement questions: (1) Visual Detail Questions (×6): Statements regarding visual details in the critical shots and (2) Event Sequence Question (×1): A statement regarding the sequential order of the critical scene in relation to other events happened in the video. The participant was instructed to indicate “True” if the item correctly described exactly what happened in the video, and “False” if the item was not correct. After answering all the questions, participants moved on to the next video. For all the participants randomly assigned to the control group, they were not shown any video, but were directly prompted to answer six sets of true-or-false judgment questions regarding visual details and event orders in the six videos.

After the participants finished all the videos, they were asked a question about their perceived difficulty of paying attention to both the sequence of actions and contextual details during the experiment. Upon completion, participants received a message saying they have completed the experiment.

Results

For the reversal detection task, participants were scored as detecting the misordered events if their written response correctly identified the item involved in the critical clips and communicated what they perceived during the reversal. Using videos without blanks in between, participants’ performance of detecting misordered events remained relatively low (38.5% correct, False Alarm = 8.6%).

Participants in the experimental group performed significantly better than the control group on both detail questions (experimental 59.2% correct, control 52.4% correct, $\chi^2(1) = 23.738$, $p < .001$) and sequence questions (experimental 70.1% correct, control 56.7% correct, $\chi^2(1) = 11.668$, $p < .001$). For the experimental group, there was no significant difference in the accuracy for detail questions between the normal trials (58.2% correct) and the reversed trials (60.2% correct, $p = .391$, $BF_{01} = 4.882$). However, the accuracy of sequence questions in normal trials (74.7% correct) was significantly higher than in the reversed trials (65.5% correct, $\chi^2(1) = 4.250$, $p = .039$, see Table 1). There was a weak correlation between participants' ability to detect reversal and their accuracy in detail questions ($r = .267$, $p = .043$), and there was no correlation between reversal detection and accuracy in sequence questions ($r = .201$, $p = .131$).

Table 1

Results of Reversal Detection and Accuracy of Visual Detail and Event Sequence Questions

		Percentage of "Yes" Response for Reversal Detection	Visual Detail Question	Sequence Question (About the critical event)	Sequence Question (Not about the critical event)
Exp. 1 (n = 58)	Normal Videos	8.6%	58.2%	*74.7%	
	Reversed Videos	38.5%	60.2%	65.5%	
Exp. 2 (n = 63)	Normal Videos	8.9%	59.6%	68.3%	77.7%
	Reversed Videos	36.5%	61.5%	72.0%	83.5%
Control (Exp.1) (n = 20)			52.4%	56.7%	

*p < .05

Discussion

Experiment 1 replicated the finding by Hymel et al. (2016) that the detection of sequence reversals was difficult even when people were instructed to pay attention to misordered events using more naturalistic videos without blanks between shots. The result of the reversal detection rate in the current experiment (38.5% correct, False Alarm = 8.6%) was comparable to and even lower than the result of the intentional reversal detection experiment in the Hymel et al. study (53.3% correct, False Alarm = 6.7%), which further strengthens the idea that reversal detection requires goal-directed attention and may not be a default process during normal event perception.

For the two memory tasks, the visual detail memory task shows that the presence of reversals in the video did not increase viewer's encoding of specific details in the reversed event, which suggests that seeing unusual temporal information did not automatically enhance viewers' sensitiveness to sensory input at the moment. This further limits the impact that temporal disruption may create on the renewal of mental event representations, because according to the event segmentation theory, the updating of event model should be implemented by temporarily amplifying the impact of perceptual input. In addition, the result of the event sequence task demonstrated that seeing reversed actions negatively affected viewers' memory of the serial location of the current event. This contradicts our original hypothesis that the presence of reversal might trigger the process of reestablishing of a more accurate higher-level representation regarding the current event. Instead, it suggests that the lower-level processing of misordered sequential information may absorb some attention from the simultaneous higher-level process of encoding the current event into the general event sequence. The minimal correlation we observed between viewers' reversal detection ability and their performance on the two memory tasks shows that the conscious detection of misordered events and the encoding of event details are likely to be processed independently.

Since the memory tasks employed in Experiment 1 revealed very little of the impact reversals created on event perception, our next step was to use event segmentation as a more direct measure to examine how viewers' experience of event boundaries might be influenced because of the misordered information. We also wanted to see if the negative effect created by reversals on encoding the serial location of the misordered event could be replicated or even generalized to other events in the video.

Experiment 2

As reviewed in the introduction, many authors have demonstrated that temporal information plays a central role in organizing people's mental representation and understanding of events. According to the Event Segmentation Theory (Zacks et al., 2007), reversing the temporal order of two events might generate an ambiguous and less predictive perceptual experience, thereby triggering the viewer to segment and experience more event boundaries. In this experiment, we asked the participants to mark the boundaries they experienced while watching each video, in an attempt to see if the number of segmentations they make increases during the reversal or after seeing it. We also implemented the same two memory tasks as in Experiment 1 to see if the previous findings replicate, and made some improvement on the event sequence memory task by adding another type of questions targeting events other than the critical event in the video. Different from Experiment 1, we did not warn the participants about the existence of event misorderings at the beginning and implemented the reversal detection task as an incidental task. The reason was that we tried to make sure participants' primary task in the experiment was event segmentation. Since the structure of the incidental detection task was similar to the task used in the study by Hymel et al. (2016), we expected the hit rate of the incidental detection task to be near-zero as in the previous study.

Participants

66 Vanderbilt University students participated in the study in exchange for course credit. Among all participants, three of them were excluded for not finishing the whole experiment, thus resulting in a total of 63 records being used in the analysis (46 female, average age = 19.0 years, SD = 1.0 years). The experiment was approved by Vanderbilt University’s institutional review board, and all participants have completed informed consent forms prior to the experiment.

Stimuli

The videos used in the experiment were the same as in Experiment 1. One additional movie showing someone dressing to go outdoors (24s) was added as the segmentation practice video in the practice phase. This movie was filmed and edited in the same way as other movies and was shown in its normal sequence.

Design

This experiment used a within-subjects design. Similar to Experiment 1, the independent variable was still the presence of misordering in each video viewed by the participants. Each participant watched and segmented six videos, three of which included reversed events. Several measurements in Experiment 1 were included again: participants’ performance on their (1) memory for visual details during the reversed event, (2) memory for the serial location of the reversed event, and (3) ability to detect reversals. Another measurement, participants’ memory for the serial location of other events was added for the event sequence task. The key measurement in Experiment 2 was the way participants segmented the videos they viewed.

Specifically, we tracked the total number of segmentations they make for each video, and the timing of each segmentation for each participant.

Procedure

Participants accessed an online experiment compiled using Quatrics. Before they began the experiment, participants were instructed to provide some demographic information, including their age and gender. Then they read an instruction about their task during the experiment. They were told that they would be watching six short silent videos that depict everyday activities. In contrast to Experiment 1, they were neither given the definition of misordering in videos nor informed of the possibility of seeing misordering in the videos. Participants were told that they would be watching each video twice in a row with different tasks: The first time they watch a video, they would need to pay close attention to the events in order to answer questions about details in the video later; The second time they watch the video, they would need to press the "N" key whenever they believe one meaningful event ends and another event begins.

Before the experimental phase started, participants entered a practice phase to practice their familiarity with the event segmentation procedure. The result in the practice trial was not included in the data analysis.

In the experimental phase, each participant watched three normal videos and three videos containing misorderings. The order of the six videos and their accompanying questions was randomized for each participant. Each video was played twice, and the participant was prompted to segment for events only in the second play. Before each video started, there was a countdown for five seconds. All videos were played at a resolution of 640 * 360 on the participants’ own digital device, and the average length of each video is 22.5s. After watching and segmenting each video, participants completed both visual detail and sequence memory tasks: They answered six questions regarding the visual details in the reversed event, one question regarding the serial location of the reversed event, and one question regarding the serial location of other events in the video.

After the participants finished all the videos, they were given the definition of a "reversal" in the video and were instructed to answer a series of questions about reversals in the video. They were first asked "Did you see any backward action while you were watching the videos?" If the answer was yes, they would report the number of backward actions they remember seeing and engage in a recognition task to identify the video they remember seeing backward actions in based on the thumbnail pictures of each video. At the end, they were asked a question about their perceived difficulty of paying attention to both the sequence of actions and contextual details during the experiment. Upon completion, participants received a message saying they have completed the experiment.

Results

Participants' ability to detect misorderings in the videos incidentally in Experiment 2 (36.5% correct, False Alarm = 8.9%) was comparable to the performance in Experiment 1 (38.5% correct, False Alarm = 8.6%). There was no significant difference in the accuracy for visual detail questions ($t(62) = -.842$, $BF_{01} = 5.164$), sequence questions about the critical event ($t(62) = -.766$, $BF_{01} = 5.474$), and sequence questions about other events ($t(62) = -1.498$, $BF_{01} = 2.514$) between the normal condition and the reversed condition. There was no correlation existing between participants' ability to detect reversal and their accuracy in visual detail questions ($r = .035$, $p = .392$), sequence questions about the critical event ($r = .080$, $p = .532$), or sequence questions about other events ($r = .097$, $p = .449$) (See Table 1).

To examine the influence of reversals on participants' frequency of segmenting events, we tracked the number of segmentations made by each participant in each video during three different time intervals (see Table 2): (1) During the critical clips (the "grab/use" and "use/grab" scene of the critical misordered item), (2) During the critical clips and the clip that followed, and (3) During the rest of the video after the critical clips appeared. During the critical misordered event, participants made significantly more segmentations in the reversed condition than in the normal condition ($t(62) = -4.092$, $p < .001$, $d = -.516$). During the critical clips and the next clip, there was no difference between the number of segmentations made for the reversed and the normal condition ($t(62) = -1.343$, $p = .184$, $BF_{01} = 3.086$). Similarly, during rest of the video after the reversed event happened, there was no difference in the number of segmentations between the reversed and the normal condition ($t(62) = .708$, $p = .482$, $BF_{01} = 5.702$).

Table 2

Average Number of Segmentations During and After the Critical Clips

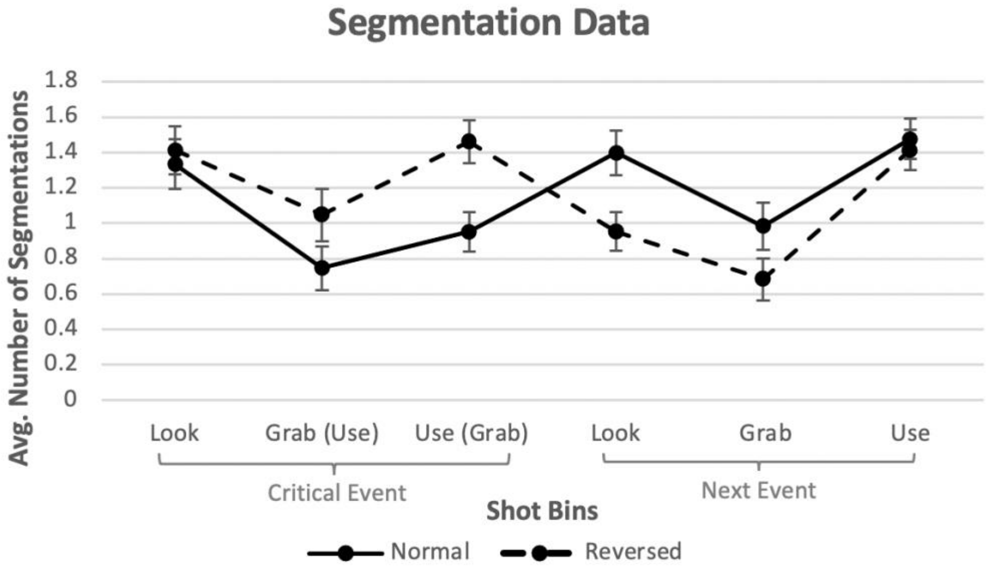
Interval	Average Number of Segmentations		Test Statistics
	Normal Videos	Reversed Videos	
Critical Clips	1.585	***2.323	$t(62) = -4.092$ $p < .001$ ($d = -.516$)
Critical Clips + Next clip	3.000	3.277	$t(62) = -1.343$ $p = .184$ ($d = -.169$) $BF_{01} = 3.086$
Critical Clips → end	8.831	8.554	$t(62) = .708$ $p = .482$ ($d = .089$) $BF_{01} = 5.702$

***$p < .001$

Based on this, we zoomed into the time interval incorporating the critical misordered event and the event next to it to look at how the average number of segmentations was distributed across each subevent. The segmentation data was binned by shots, and each event included three subevents: “Look,” “Grab,” and “Use” (See Figure 2). For the reversed condition, the peak value of segmentations occurred during the “Grab” subevent of the critical misordered event (Critical Event – Use/Grab), and for the normal condition, the peak value of segmentations occurred during the “Look” subevent of the event next to the critical misordered event (Next Event – Look).

Figure 2

Binned Segmentation Data for the Critical Misordered Event and the Next Event



Discussion

This experiment shows that the existence of sequence reversal did not increase the number of event boundaries experienced by the participant. Even though the primary analysis revealed that the number of segmentations made during the critical clips (Grab/Use and Use/Grab) was significantly higher in the reversed condition, a further analysis that broke down the segmentation data into different subevent bins revealed that this result indicated a different timing of the boundaries experienced by the viewer, instead of an increase in the number of boundaries. Looking at Figure 2, we discovered that the reason why the interval of the critical event (Critical Event – Grab/Use and Critical Event – Use/Grab) included more segmentations in the reversed condition was that the peak of segmentation in this condition (Critical Event – Use/Grab) was advanced compared to the normal condition (Next Event – Look). There are two possible explanations for this result: Viewers in the reversed condition treated either (1) the signature subevent signaling the end of an event (i.e., the “use” subevent itself), or (2) the reversal (i.e., the abnormal return from “use” to “grab”) as a cue for segmentation. In either case, the result was that their segmentation landed in the next subevent (Critical Event – Use/Grab) after seeing the signal. In contrast, viewers in the normal condition tended to treat the natural event boundary between two events (the transition from the “Use” clip of the current item to the “Look” clip of the next item) as a cue for segmentation. Therefore, their segmentation landed in the next subevent (Next Event – Look) after the natural transition. Taking this factor into consideration, we concluded that the presence of reversal did not increase the number of segmentations made during or after the reversal, but the unusual visual information produced by the reversal itself did induce a subjective experience of an event boundary. It is worth noting that after viewers in the reversed condition segmented in response to the reversal, they failed to respond to the more regular event boundary between the current item and the next item, thereby resulting in the same overall number of segmentations as in the normal condition.

As for other measurements besides event segmentation, Experiment 2 yielded a rather surprising result that the performance in the incidental reversal detection task was almost identical to the result of the intentional reversal detection task in Experiment 1, and was a lot better than the near-zero level performance in the incidental conditions in the Hymel et al. (2016) study. This suggests that the task of segmenting events might increase people’s awareness of sequence reversals, and we discussed this possibility in further detail in the General Discussion section.

Experiment 2 replicated the finding in Experiment 1 that the presence of reversals did not enhance visual detail encoding during critical clips, but the negative impact caused by reversals on the overall event sequence memory disappeared. Similar to the finding in the previous experiment, Experiment 2 revealed no correlation between viewers’ ability to consciously detect reversals and their performance in answering questions related to visual details or event sequence memory.

General Discussion

In the current study, we use possibly more sensitive behavioral measures to reveal the impact of sequence reversal on people’s perception and representation of events. Experiment 1 replicated the finding of Hymel et al. (2016) that reversal detection was difficult for viewers even when they were told explicitly to detect it, which further strengthens the idea that the awareness of sequence reversal is not likely to be a default process during event viewing. Experiment 2 indicated a new possibility that when being asked to perform specific tasks during viewing, viewers’ performance in incidental reversal detection could be increased to a level that was comparable to intentional detection tasks in Experiment 1. In addition, Experiment 1 demonstrated that the existence of reversed action sequences did not increase viewers’ encoding of visual details during the reversal. However, these added tasks negatively impacted viewers’ ability to remember the serial location of the current event in the whole event sequence. Experiment 2 demonstrated that the presence of reversals did not increase the number of event boundaries experienced by the viewers during or after the reversed event. However, a closer analysis revealed that the reversals changed the timing of the event boundary experienced by the viewer: The unusual visual information created by the presence of the reversal induced a subjective experience of an event boundary, and subsequently prevented the viewer from perceiving the more regular type of event boundary mediated by item differences that came after the reversal. In addition, the effect that the existence of reversal impaired viewers' ability to encode event sequence information was not observed in Experiment 2.

Possible Implications of the Event Segmentation Task

In Experiment 2, one surprising finding on reversal detection was that with the presence of an event segmentation task, the level of performance for the incidental detection of misordered events was almost similar to the intentional detection in Experiment 1. This result was very different from the near-zero performance in the study by Hymel et al. (2016), even though the design of these two incidental detection experiments shared many similarities: In the previous experiment, participants also watched multiple (12) videos about familiar everyday events and performed a primary task (detecting the critical misordered clip) for each video. Similar to Experiment 2 in the current study, participants were not notified about the presence of the reversals at the beginning of the experimental phase and were asked whether they noticed any misordering in the videos after they have completed all the trials. Therefore, the discrepancy between the result of these two experiments might lead us to ask what were the factors that caused this notable increase in incidental reversal detection ability. One possible explanation is that the repeated viewing of the same videos in Experiment 2 increased participants' familiarity with each video. The fact that each participant viewed each video once before they started to segment the event in the second viewing gave participants more opportunities to spot the reversal. Another explanation for the increase in performance would be due to the particular nature of the event segmentation task in Experiment 2: The action of consciously marking the boundary between different events in the video helped participants direct more attention towards the temporal structure of events, thus allowed them to be more aware of the presence of sequence order violations. Based on this finding, it will be particularly interesting to run a follow-up experiment and test whether it is repeated viewing or event segmentation that can create a boost effect in incidental reversal detection.

In addition, if we compare Experiment 2 with Experiment 1, one major difference in memory task performance was that the inhibitory effect produced by reversal on people's ability to remember event sequence in Experiment 1 disappeared in Experiment 2. Similar to the above discussion, it would also be interesting to conduct a follow-up experiment to examine whether it was repeated viewing or event segmentation that caused this effect. If the result favors event segmentation, it may indicate that paying more attention to the beginning and ending of each event helped viewers to become more robust about the general structure of the event, and therefore their performance is less likely to be influenced by the presence of the reversals.

Possible Mechanisms of Processing Sequence Reversals

The result of the current study extends our understanding of how misorderings in action sequences are processed during dynamic event perception: Despite the fact that these reversals frequently escaped viewers' conscious detection, the timing of the information causing the reversal did not completely fail to enter viewers' perceptual processing stream. According to the event segmentation theory (Zacks et al. 2007), the fact that this information elicited an event segmentation on the behavioral level signifies that it should have generated a spike of prediction errors and caused the updating of some current event representation. However, it is apparent that this process did not guarantee the reversal itself to enter viewers' awareness, and the process of the current model updating that elicited the boundary did not necessarily lead to a large-scale event model reconstruction whose enhancement can be captured by our measures of visual event details and event sequence memory.

There can be two possibilities regarding the specific type of information created by the reversal that leads to an advance in the timing of event segmentation: One is the advanced signature subevent itself (e.g. Seeing the salmon being put into the bowl in the "Use" clip earlier leads the viewer to believe that all the actions about the salmon have been finished, and that the actions on the next item should occur subsequently), and the other is the reversal of the two subevents in critical clips (e.g. Seeing the abnormal return from the "Use" clip to the "Grab" clip leads the viewer to "false alarm" for an event boundary). If the latter case holds true, this phenomenon may lead us to think about the nature of how the error signals generated by the reversals are processed by our brain, and how this process is related to the broader event representation mechanism that governs our understanding of the whole event sequence. Previous theories of event cognition such as the Event Horizon Model (Radvansky & Zacks, 2017) emphasized the fact that events are represented hierarchically on multiple different time scales, and that the "current [event] model" can provide feedback to "a hierarchy of model ranging from short to longer durations" (p. 134). The Event Segmentation Theory outlined the principle for error-signal handling across different time scales (Zacks et al., 2017): Fine-grained event representations can update themselves in response to "small, brief increases in prediction error" (p. 276). In contrast, coarse-grained representations can only be reset in response to "longer, more sustained increases in error." From the result of the current study, we can infer that the reversals we created by switching the "using" and "grabbing" actions on an item can be considered as inducing brief increases in prediction error on a fine-grained scale. Even though they caused a behavioral response of event segmentation and an internal updating of the fine-grained event model, they failed to impact the coarse-grained representation that is more closely related to viewers' conscious awareness and their long-term memory of event contents.

It is also important to consider what might be the exact mechanism that prevents certain updating on fine-grained event models to enter the awareness and reach the threshold of altering the representations in coarse-grained event models. A core component in the Event Segmentation Theory that is particularly relevant to explaining the phenomenon in the current experiment is the top-down influence from Event Schemata on Event Models. Viewers often acquire their event schemata from previous life experiences, which contains important information about the sequential structure of activities that can be used to guide their processing of newly encountered events (Zacks et al., 2007). In the scenario of the current study, our viewers are likely to bring their schemata of how certain actions constrain each other into ordered sequences in real life into the viewing of our video stimuli: When the reversal between "grab" and "use" occurred, viewers' fine-grained model representing the temporal relationship between these actions were updated and sent out some error signals for event segmentation. However, their event schemata containing the information that "the most common sequence of taking an item is look → grab → use" exerts a strong top-down influence and prevented the coarse-grained event model from accepting the input from the perceptual stream. This explanation strongly suggests the existence of an event-processing system that can dynamically allocate cognitive resources to enable the most efficient encoding of event sequences. As was argued by Hymel et al. (2016), when the benefit of engaging in moment-to-moment prediction and detecting error-signals fails to outweigh the cost of devoting more cognitive resources, the event schemata may exert more influence and cause some temporal information of the event to be left outside of awareness. This causes insensitivity to fine-grained temporal errors.

In the future, it might be interesting to replicate the current study using video stimuli that contain other types of temporal violations besides the reversal between “use” and “grab.” In particular, it might be useful to create scenarios where detecting sequence reversals are necessary for viewers’ to effectively understand the higher-level meaning of the whole event, to see if goal-directed attention will drive the perception of a larger proportion of temporal errors to enter the conscious processing stream.

References

Aslin, R. N, Saffran, J. R, & Newport. E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science, 9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Baker, L. J., & Levin, D. T. (2015). The role of relational triggers in event perception. Cognition, 136, 14–29. https://doi.org/10.1016/j.cognition.2014.11.030Brogaard, B. (2011). Are there unconscious perceptual processes? *Consciousness and Cognition*, *20*(2), 449–463. https://doi.org/10.1016/j.concog.2010.10.002

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. https://doi.org/10.1037/0278-7393.32.5.1031

Creel, S., Newport, E., & Aslin, R. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(5), 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, *22*(2), 243–252. https://doi.org/10.1177/0956797610393742

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499–504.

Fiser, J. Z., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.

Fiser, J. Z., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA, 99*, 15822–15826.

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11*(12), 10.1167/11.12.3 3. https://doi.org/10.1167/11.12.3

Hymel, A., Levin, D. T., & Baker, L. J. (2016). Default processing of event sequences. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 235–246. https://doi.org/10.1037/xhp0000082

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition, 39*(1), 75–91. https://doi.org/10.3758/s13421-010-0027-2

Kurby, C. A., & Zacks, J. M. (2021). Priming of movie content is modulated by event boundaries. *Journal of experimental psychology. Learning, memory, and cognition*, 10.1037/xlm0001085. https://doi.org/10.1037/xlm0001085

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220–244.

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

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences, 17*, 133–140. https://doi.org/10.1016/j.cobeha.2017.08.006

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), 1–8. https://doi.org/10.1016/j.ijpsycho.2010.01.006

Richmond, L. L., Gold, D. A., & Zacks, J. M. (2017). Event perception: Translations and applications. *Journal of Applied Research in Memory and Cognition*, *6*(2), 111–120. https://doi.org/10.1016/j.jarmac.2016.11.002

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, *18*(1), 37–55. https://doi.org/10.1002/acp.940

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology*, *21*(22), R912–R913. https://doi.org/10.1016/j.cub.2011.09.049

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*(1), 118–132. https://doi.org/10.1016/j.cognition.2009.12.003

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

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., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the alzheimer type. *Psychology and Aging*, *21*(3), 466-482. http://dx.doi.org /10.1037/0882-7974.21.3.466