THE DETECTION OF UNEXPECTED EVENTS AND IMPLICATIONS FOR EVENT PERCEPTION

By

Alicia Marie Hymel

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Psychology

August, 2013

Nashville, Tennessee

Approved:	Date:
Daniel Levin	07/17/2013
Megan Saylor	07/17/2013

ACKNOWLEDGEMENTS

This thesis is based upon work supported by the National Science Foundation under grant #0826701 “Thinking About, and Interacting with Living and Mechanical Agents,” awarded to Dr. Daniel Levin (PI), Dr. Gautam Biswas, Dr. Julie Adams, and Dr. Megan Saylor.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

A key process underlying event perception is the segmentation of continuous visual input into the discrete actions that make up an event. For example, the event "making a cup of coffee" could be segmented into pouring the coffee into a cup, adding sugar and milk, and stirring the coffee. It has been proposed that individuals segment events by engaging in continuous prediction contingent upon sequential regularities, with event boundaries occurring when these predictions fail (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). More generally, the existence of predictive mechanisms for facilitating perception has long been discussed in the vision science literature. However, existing research may overemphasize the importance of moment-to-moment conceptual predictions in event segmentation and perception. In particular, it is possible that automatic predictions corresponding to higher-level goals and sequence structure are not able to guide the ongoing segmentation of events. Consequently, naturalistic perception of a coherent action sequence may be influenced by some non-automatic strategic capacity-limited predictive process, rather than these proposed default predictions. The goal of the studies presented here is to determine the extent to which individuals are able to use these strategic predictions in real time event perception.

Researchers have explored different ways that prediction might be used in perceiving and understanding events. At a basic perceptual level, future locations of objects are extrapolated based on prior motion. This effect has been studied in the context

of representational momentum. Freyd and Finke (1984) showed participants a stationary rectangular stimulus that rotated in an implicit 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 implicit direction of motion than had been previously shown. The participants appeared to be engaging in a low-level prediction of the future location of the stimulus.

Research on reaching and object tracking also provides evidence for how prediction and action can interact with prior knowledge. Von Hofsten, et al. (1998) presented infants with a moving object that could either travel in a straight line across a surface, or change directions at the midpoint of its movement trajectory. Infants' patterns of head turning and reaching suggested they predicted that the object would move along a linear path. Even on trials in which the movement was non-linear, the infants' heads continued to move for approximately 200 ms after the object changed its path. Additionally, on non-linear trials, infants' hands moved towards the area of space where the object would have traveled if it had moved linearly. Even at an early age, simple concepts about object motion, such as the idea that an object will not change its direction of motion unless acted upon by an outside force, are able to guide perception and prediction.

However, prediction is not necessarily limited to simple forward projections of motion or location extrapolation based on fundamentals of Newtonian physics. There is some evidence to suggest that concepts can affect these lower-level processes. 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.

For example, an object that typically moves great distances, such as a rocket, will elicit more representational momentum, i.e., will be displaced further along its implicit direction of motion, than an object that is typically stationary, such as a building. Prior knowledge and concepts beyond applications of basic physics affect predictions about future states.

The use of predictions in both perception and action can be even more conceptually oriented than suggested by motion extrapolation alone. Theory of mind is the process by which individuals reason about the mental states of others, including their intentions and goals (Premack & Woodruff, 1978). This skill is an important component of our ability to predict others' behavior. While the prediction occurring in representational momentum and even reaching appears to be automatic and effortless, predictions based on beliefs about others' mental states may not be. For example, this may be the case with the attribution of false beliefs. In one study, a false-belief task was used to test whether belief inferences are automatic (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006). Participants watched a video in which there was a male and a female agent. Both agents saw an object placed in one of two containers. Afterwards, the female agent left the room, and the location of the two containers was switched. The female agent now falsely believes that the object is still in the previous location. Subjects were then presented with either a belief probe (e.g., "She thinks that it's in the box on the left.") or a reality probe (e.g., "It is true that it's in the box on the right."). Participants were slower to respond to the belief probe than the reality probe. Additionally, if subjects were explicitly instructed to track the female agent's mental state, there was no difference in reaction time for the belief and reality probe responses.

The authors argued that this pattern of results indicated that while information regarding reality is automatically processed, information about others' beliefs is not. Therefore, it appears that any predictions based upon an individual's beliefs, and perhaps, by extension, their desires and goals, may require time and effort to produce.

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, to what extent is prediction useful in perceiving and understanding the unfolding of familiar events over time? One proposed use of prediction in event perception is determining when one event ends and another begins, or event segmentation. Event segmentation is the process by which continuous action is partitioned into meaningful events in real time. In laboratory conditions, event segmentation is typically measured by asking participants to watch a video and make a response when one meaningful event ends and another begins. If participants are asked to segment videos into the smallest (fine-grained segmentation) and largest (coarse-grained segmentation) units that seem natural, they appear to do so in a way that is hierarchical, as coarse event boundaries tend to be aligned with fine event boundaries (Zacks, Tversky, & Iyer, 2001).

The temporal placement of event boundaries has been shown to correlate with perceptual change, such as movement. Using human actors and a dance notation measuring changes in body position, one study showed that event boundaries coincided with a greater number of physical changes than intervals between the boundaries (Newtson, Engquist, & Bois, 1977). Similar studies have been conducted using computer-generated stimuli. In a study asking participants to segment a video of

animated geometric figures similar to that shown in Heider and Simmel's (1944) classic experiment, both coarse and fine boundaries corresponded to increases and changes in movement (Hard, Tversky & Lang, 2006).

In addition to these basic perceptual characteristics, it has been argued that event segmentation is affected by conceptual changes and knowledge. However, this evidence is less direct. One experiment showed that tones inserted near goal completions were better remembered than tones inserted at other locations within a video, and the tones that did not occur near the goal completions were remembered as occurring nearer to goal completions than originally presented (Baird & Baldwin, 2001). The authors argued that this pattern of results implies that perceivers group activity into units based on inferred intentions, as images taken from movie event boundaries are better remembered than those from non-boundary moments (Newtson & Engquist, 1976).

Another study utilized videos of animated objects participants believed either to have been generated by people controlling the objects as in a video game, or via computer-generated random movement trajectories (Zacks, 2004). The relationship between stimulus movement and event segmentation was stronger in fine segmentation than in coarse segmentation, suggesting that non-movement features, perhaps including inferred goal acquisition, may affect coarse-grained segmentation. Additionally, the relationship between movement and event segmentation was stronger when participants believed the movie portrayed randomly generated motion, rather than goal-directed behavior. The authors argued that these results indicate that prior knowledge about goal states and goal acquisition affects how physical cues are processed when identifying event boundaries. However, whether participants used information about potential goal

states remains unclear. Additionally, if the participants were using information about goal states when generating event boundaries, the sophistication of these goals is difficult to elucidate. For example, one can conceive of a goal of moving to a particular location in space. However, a more complex goal may organize a series of discrete motions occurring in a particular order.

While event boundaries have been shown to correlate with both perceptual and, perhaps, conceptual aspects of ongoing activity, such findings do not explain why event boundaries are placed where they are. One possibility is that boundary placement is based on predictions about future events. Event Segmentation Theory (Zacks et al., 2007) relies upon the continuous generation of perceptual predictions as an explanatory mechanism for segmentation. 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 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). 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. The resulting increase in

prediction error causes the updating of the event model, which is perceived as an event boundary.

While the use of predictive mechanisms in visual perception has received much attention from researchers, we were specifically interested in examining the use of predictions about future events in event perception. Event segmentation does not necessitate the use of a continuous predictive mechanism and the detection of prediction errors as the impetus for segmentation. Instead, segmentation could be a concomitant of one or many perceptual and conceptual changes, without an explicit reliance on prediction failure. Alternatively, one implication of hierarchical theories of event perception is that actions may be processed in terms of their links to higher-order goals. Consequently, event perception and segmentation may rely on deviations from these broad overarching representations but involve relatively little moment-to-moment prediction. Similar to the theory of mind findings presented in Apperly et al., 2006, processing the actions that make up an event may be automatic, but determining how these events compare to our predictions may require additional processing. To investigate this possibility, we conducted a series of experiments testing the extent to which event perception involves this type of conceptual prediction. 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. Importantly, the

stimuli used here allow for generalizability beyond many past studies as they depict everyday events and do not involve repetitive actions.

CHAPTER II

EXPERIMENT 1

In Experiment 1, we were interested in obtaining base rates of misordering detection (and false alarms) for our videos. Our hypothesis was that if making predictions about upcoming actions is important in event perception, then the detection of out of order actions should be facilitated, as these actions should not have been predicted. If these predictions are not continuously being made, however, detection would not be facilitated, as there would be no on-line predictions available to use when comparing perceived and expected events.

We also examined the effects of a secondary verbal task on performance. If these conceptual predictions are used in perceiving and understanding events in our daily lives, then it can be assumed (and is assumed by models such as Event Perception Theory) that this process is continuous and automatic. If this is true, then we hypothesized that an interfering task would have no effect on the detection of out of order actions.

In these videos, the out of order action was always located one shot later in the action sequence than it would have otherwise belonged. For example, one of the twelve videos contained an out of order action sequence showing an actor stirring a cup of coffee with a stirrer, after which she picked up the same stirrer from a table and then removed the stirrer from the cup. This can be compared to the correct action sequence of picking up the stirrer from the table, stirring the coffee, and removing the stirrer from the cup. A complete list of the actions contained in this particular stimulus video can be found in

Figure 1, with the misordered action occurring in the tenth clip. This clip sequence ensured that not only did these out of order actions appear unlikely, but also impossible without some elaborate narrative construction on the part of the participants (e.g., "Perhaps the video did not show the actor removing the spoon from the cup after stirring it, and it then showed the actor stirring the coffee a second time with a new spoon.").

Clip	Action Description	
	Normal Video	Misordered Video
1	Establishing shot of actor near coffee pot	Same as Normal Video
2	Actor picks up coffee pot	Same as Normal Video
3	Actor pours coffee into a cup	Same as Normal Video
4	Actor puts coffee pot in its prior location	Same as Normal Video
5	Actor picks up creamer	Same as Normal Video
6	Actor opens creamer container	Same as Normal Video
7	Actor pours creamer into coffee cup	Same as Normal Video
8	Actor opens sugar container	Same as Normal Video
9	Actor pours sugar into coffee cup	Same as Normal Video
10	Actor removes stirrer from its holder	Actor stirs coffee with stirrer
11	Actor stirs coffee with stirrer	Actor removes stirrer from its holder
12	Actor shakes excess liquid off stirrer	Same as Normal Video
13	Actor picks up coffee cup	Same as Normal Video
14	Actor drinks from coffee cup	Same as Normal Video

Figure 1. Example Action Sequence in Normal and Misordered Videos

Participants

Fifteen undergraduate students (13 female; mean age = 18.80 years, *SD* = .68) were recruited from Vanderbilt University. Participants were compensated with course credit.

Methods

All videos were played at a resolution of 740 x 480 on a 17-inch screen with a 1280x1040 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 twelve video sequences consisting of an actor performing a familiar task (e.g., making a cup of coffee, using a copy machine). The sequence was filmed using multiple camera angles, and shots were zoomed in and centered on the most relevant actions after an initial establishing shot. Each shot within a video sequence was separated by 67 ms of blank screen. These angle changes, or cuts, segmented the sequences into a set of discernible sub-actions (e.g., stirring a cup of coffee, lifting the lid of a copy machine). Additionally, some clips in these videos were slightly sped up in order to keep clips short. One representative examples of this was a shot of a girl walking across a room in a video about sharpening a pencil. The average length of the misordered shots was 533 ms, the average length of the videos was 9583 ms, and there was an average of 11.25 shots within the videos. For all participants, half of the videos showed the events in the correct order, while the other half contained a misordering in which the order of two actions was switched. 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.

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 were would be one misordering in which two shots would be out of order. 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 backwards from a given number by three. Before these trials, this number was displayed on the screen for 5000 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 across all of the participants, for any one video, four equally represented trial types existed: misordering and a counting task, correct order and counting task, misordering and no counting task, and correct order and no counting task.

Results

Analyses were conducted to test whether correct classification of the videos was affected by the presence of a secondary verbal task (interference, no interference) for

both video types (misordering, no misordering). For videos containing a misordered event, participants were significantly more likely to detect the presence of the misordering on trials without a secondary task (mean = 53.33% correct, *SD* = 27.60) compared to trials with a secondary task (mean = 24.44% correct, *SD* = 26.62; $t(14) = -2.83$, $p = .01$). However, for videos without a misordered action, the presence of a secondary task had no effect on main task performance. Participants were able to correctly reject the presence of a misordering for 93.33% (*SD* = 13.80) of trials when there was no secondary task and 86.67% (*SD* = 21.08) of trials when there was a secondary task ($t(14) = -.90$, $p = .38$).

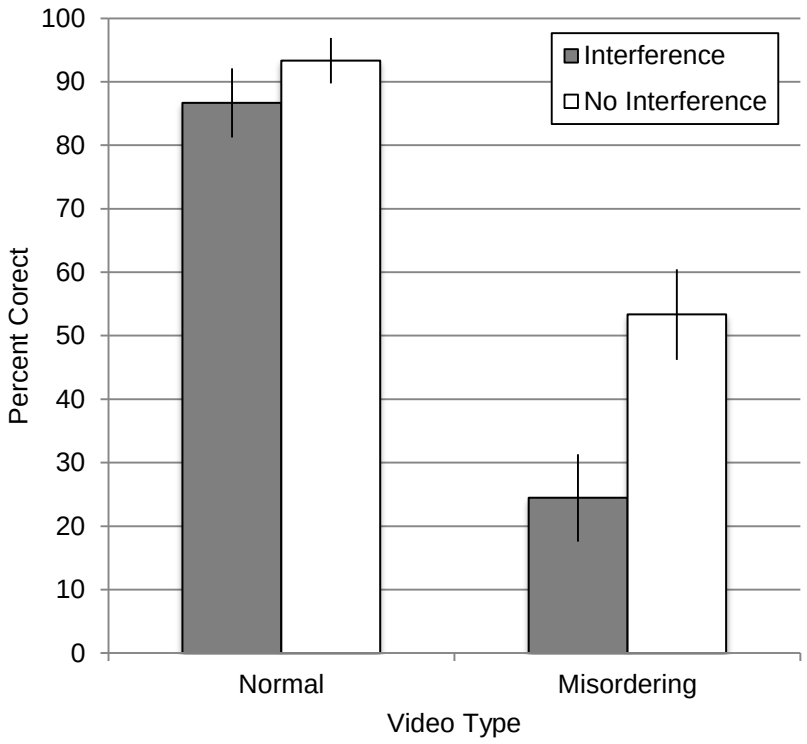

Figure 2. Effect of Video Type and Secondary Task on Correct Video Classification

We also compared the number of verbal counting responses made across trials that ended in a correct or incorrect identification of video type (misordering or no misordering) to determine whether performance on the secondary task differed as a function of accuracy on the main task. For trials containing the secondary verbal task, participants gave a verbal count an average of every 1.61 seconds (*SD* = .48) for trials ending in a correct response and every 1.63 seconds (*SD* = .72) for trials ending in an incorrect response. There was no significant difference in the number of secondary task responses per second between trials ending in a correct or an incorrect primary task response ($t(14) = .14$, $p = .89$).

Discussion

In Experiment 1, we determined that not only is the detection of unpredictable, out of order events difficult for participants, but their performance on this task further deteriorates when asked to complete a secondary task. This inability to reliably detect event misorderings indicates that perception may not involve the moment-to-moment tracking required in order to predict actions on the timescales represented by these stimuli, or that participants may have difficulty accessing the content of any such predictions for use in self-report paradigms. Also, the addition of a secondary task significantly lowered detection, indicating that generating or using these predictions in event perception is not purely a bottom-up mechanism, and online prediction generation may deteriorate in real world scenarios under divided attention.

Next, we investigated rates of incidental misordering detection. In Experiment 1, participants expected to see these out of order actions and were explicitly asked to look

for them. While participants had difficulty detecting the out of order sequences, we were interested in whether their performance would deteriorate further if the misordering detection task was incidental. This incidental version has the additional benefit of being more similar to typical event perception. If prediction of upcoming events is important in event perception, participants should be able to complete this task without deliberately trying to detect when an unexpected event occurs.

CHAPTER III

EXPERIMENT 2

Participants

Twelve participants (8 female; mean age = 23.41 years, *SD* = 6.08) were recruited from Vanderbilt University and the surrounding community. Participants were monetarily compensated.

Methods

All videos were played at a resolution of 740 x 480 on a 13-inch laptop screen with a 1280x800 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. They were not informed of the presence of the misordering. Every participant only watched one of the original twelve misordered videos, with each video being watched once across all the participants. Immediately after watching the video, they 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

Each of the twelve participants failed to detect the out of order action. No 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. While some participants reported noticing something usual about the video, in follow-up questioning, this always directly related to the video's editing (for example, the presence of the black screen between individual shots).

Discussion

Experiment 2 provides evidence for a lack of incidental detection of misordered actions. One issue with Experiment 1 is that participants were asked to detect these misorderings. It is possible that this request leads them to engage in additional processing or strategies to complete the detection task. In Experiment 1, the secondary verbal task should have prevented the implementation of some of these strategies, such as on-line generation of a list of scenes that the participant could analyze upon the video's conclusion. Experiment 2's results serve to strengthen the hypothesis that people do not predict future actions, at least at this timescale, in order to perceive and understand events.

One possible explanation for the results of both Experiments 1 and 2 is that the videos were difficult to understand. Predictions about future actions, based on preexisting event models and schemata, may not be generated if an event is not understood, as a usable event schema upon which to base these predictions may not be available. The short length of each clip, along with the brief black screen inserted between each action,

is atypical when compared to the way in which events are usually perceived. We addressed these concerns in Experiments 3a and 3b.

CHAPTER IV

EXPERIMENT 3

In Experiment 3, we tested the degree to which participants may have had difficulty perceiving and understanding the events presented to them in Experiments 1 and 2. Two potential problems were investigated: whether participants had difficulty understanding what was taking place in each video as a whole, and whether participants had difficulty perceiving the actions taking place in individual clips.

Experiment 3a

In Experiment 3a, we tested participants' understanding of each video sequence. If participants do not understand what is happening in a particular video, then it would seem reasonable that they could not generate predictions for future events, as they would have no event schema, or perhaps an incorrect event schema, from which to draw their predictions. New versions of each of the twelve videos were created, and participants' understanding of these videos was compared to their understanding of the original video stimuli. In particular, we wanted to determine whether the atypical editing style and sped up actions of the original videos negatively influenced participants' understanding of the events.

Participants

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

Methods

Stimuli were presented on the same equipment as Experiment 2. Re-edited versions of the twelve 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 msec black screen between each action was removed. Second, the videos were re-edited 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 seconds (*SD* = 10.00).

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 twelve videos as had been shown in Experiment 1. The other half of the participants saw the new, re-edited 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 (2 to 6 actions per video). The rater indicated whether the participant reported 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 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; $t(31) = 2.29$, $p = .03$). While this difference was significant, participants still reported an overwhelming majority of actions, regardless of video type. While participants generally performed better in the typically edited video condition, average performance was better in the atypically edited video condition for four of the twelve videos.

We were interested in determining whether misordering detection performance was poorer in our previous experiments for the videos that showed a relatively lower recall performance in the atypically edited video condition. Misordering detection performance for each individual video was compiled from past experiments. These experiments include select trials from Experiment 1, along with four additional experiments not included in this manuscript. Only trials: 1) with an out-of-order action, 2) using the original video stimuli with no extra tasks (mimicking the no-interference trials of Experiment 1), and 3) using the same population and presentation methods as the

experiments described here, were considered. Using these criteria, we were left with 309 total responses, with 24 to 29 responses per video. Across the twelve 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.

Experiment 3b

In Experiment 3b, we tested participants' ability to detect the critical out of order clip itself in each video. This "critical clip" is the action that participants must see in order to detect the misordering. For example, in the action sequence: stir coffee with a spoon, pick up spoon from table, and remove spoon from coffee, it is necessary that participants see the actor picking up the spoon from the table in order to determine that the action sequence was out of order. Even if participants have an understanding of the main event taking place in the video (as tested in Experiment 3a), an inability to recognize this particular clip would still lead to decreased performance on a misordering detection task. If participants do not detect this clip, it would appear that perhaps an action is missing, but the sequence appears in the proper order. In this experiment, we also continued to test participants' incidental detection of misordered action sequences via a brief post-test questionnaire.

Participants

Nineteen participants (11 female; mean age = 20.26 years, *SD* = 4.47) were recruited from Vanderbilt University and the surrounding area. Participants were compensated with candy, course credit, or cash.

Methods

Participants watched the same twelve videos shown in Experiment 1, and the videos were presented using the same equipment as Experiment 2. Each video was preceded by the critical target clip that was the second of the two actions involved in the misordered act, as described above. Figure 3 contains a partial list of actions in the coffee making video, with the critical clip indicated in bold. In this particular video, the critical out of order action is "Actor removes stirrer from its holder", which is the 11^{th} clip of the misordered video. The same target clip was used in videos without misordered events, although the clip appeared one action earlier in these videos due to the lack of misordering. In this example, "Actor removes stirrer from its holder" is the 10^{th} clip in the normal version of the video.

Each trial began with a 5000 ms countdown, followed by a 1000 ms black screen and the critical target 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 target clip appear in the subsequent video. Reaction times were measured from target onset. Reactions were recorded with Final Cut Pro, which is able to flag a frame of a video when a key is pressed. We were then able to compute the number of frames between this flagged frame

and the first frame of the target clip. Reaction times were generated by converting this number of frames into milliseconds. If multiple responses were made during one trial, the latest RT was selected for analysis. Participants were not told that any of the videos contained an out of order action sequence.

Clip	Action Description	
	Normal Video	Misordered Video
…	…	…
9	Actor pours sugar into coffee cup	Same as Normal Video
10	**Actor removes stirrer from its holder**	Actor stirs coffee with stirrer
11	Actor stirs coffee with stirrer	**Actor removes stirrer from its holder**
12	Actor shakes excess liquid off stirrer	Same as Normal Video
…	…	…

Figure 3. Example Location of Critical Misordered Clip in Video Stimuli

After completing the twelve trials, participants completed the post-test questionnaire used in Experiment 2 to assess implicit misordering detection. As the questionnaire was not administered until the end of the experiment, participants were asked once about all twelve videos, rather than once for each individual video. Because of this, a participant only needed to detect one of the six misorderings across the twelve trials to be considered as having detected a misordering. Additional data, such as number of misorderings incidentally detected across the twelve trials, is not reported, as this number would likely be skewed by recall failure.

Results

Across all participants, mean target clip detection was 90.35% (*SD* = 9.32). Detection of target clips in videos containing a misordering (mean = 84.21%, *SD* = 17.10) was significantly lower than detection in videos without a misordering (mean = 96.49%, *SD* = 6.98, $t(18) = 2.93$, $p = .01$). Only 31.58% of participants reported noticing any misordered actions.

A second analysis was conducted on the types of errors present in each video type (i.e., misses). Twenty-two errors occurred across 228 trials. Three of these errors were classified as "delayed" post hoc (both RTs > 3000 ms). Two errors received a "miss" classification due to a complete lack of response. Seventeen errors were classified as "anticipations", which included responses before target initiation or under 200 ms after target onset (highest RT classified as an anticipation using these criteria was 66 ms). All anticipation errors occurred when subjects were responding to videos containing an out of order action. 76.47% of anticipation errors occurred in participants who did not notice any misorderings, with 50% of participants who did not notice misorderings producing at least one of these erroneous responses.

Discussion

In Experiment 3a, participants were able to accurately recall an overwhelming majority of the actions in the event videos. Additionally, there was no relationship between performance on this memory task and ability to detect misorderings across the twelve videos. It seems unlikely that the results of Experiments 1 and 2 were due to participants not understanding the series of actions in each video. If participants would

have displayed a consistent difficulty in reporting the actions portrayed in these videos, then it follows that their misunderstanding of the action sequences led them to fail to create an event schema, which would have hindered their ability to predict upcoming actions. While their recall was not perfect, this may be due to memory effects as well as selective reporting effects. Even though participants were told to prioritize reporting the actors' actions in the videos, it is possible that they sometimes chose not to report a particular action in favor of brevity. Additionally, upon closer examination of participants' responses, sub-100% scores were always due to a lack of information (i.e., failing to report a particular action), rather than a misrepresentation of the actions in the video (i.e., reporting an action that was not present).

In Experiment 3b, most of the target clips, which must be perceived in order to detect misordered events, were detected. It appears unlikely that the results of Experiment 1 were driven by an inability to detect the critical out of order clip. While detection rates in normal videos were higher than in videos containing a misordering, the effect was completely driven by the unique appearance of anticipation errors in responses to the out of order videos.

Experiment 3b also provided additional evidence for a lack of incidental detection, as under a third of participants reported noticing at least one of the six misorderings after watching all twelve videos. However, these particular data should be interpreted with some caution, as participants were only asked about the presence of a misordered action after watching all twelve videos. It is possible that the rate of detection may have been influenced by other factors, such as memory effects. However, these results are unsurprising considering the low rate of incidental detection also found in Experiment 2.

Interestingly, in the clip detection task, we saw a number of anticipatory responses (anticipation errors), which only occurred in videos containing an out of order event. It is important to note that for these videos, if a participant made an anticipation error, they were endorsing seeing the target clip where it would have been shown if the action sequence was in the correct order. To revisit the coffee example, if the target clip was picking up a spoon from a table, and the misordered action sequence showed an actor stirring the coffee with the spoon, picking up the spoon from the table, and removing the spoon from the cup, these anticipatory responses occurred when the coffee was being stirred. As the majority of these errors occurred in participants who did not notice any misorderings, this pattern of results may actually imply a conceptual prediction for which perceptual verification is not necessary. They may have naturally generated some on-line prediction, but did not compare this prediction to the next perceived event, and did not verify whether or not the perceived event met their expectations. Further research would be needed to fully understand if and when these "unused" predictions are generated, and what purpose they serve.

While participants appear to understand the content of the stimuli videos, another potential reason for participants' difficulty in detecting these misordered actions is that they are unable to retrieve the critical misordered clip from memory. This was tested in Experiment 4.

CHAPTER V

EXPERIMENT 4

In this experiment, we tested whether participants' inability to detect misordered events was the result of an inability to remember the critical misordered action. If ending the video immediately after the critical misordered clip increases misordering detection, then this could provide further evidence that the ability to use conceptual prediction in event perception at this timescale is an effortful process rather than some continuous on-line predictive mechanism. This argument would be even more compelling if memory for the out-of-order clip itself remained the same across videos with normal and out-of-order events, as the misordering detection results could not be explained by lack of memory for this clip. To this end, in addition to measuring misordering detection, we also tested participants' memory for the misordered clip immediately after each trial.

Participants

Fifteen undergraduate students (13 female; mean age = 20.13 years, *SD* = 1.41) were recruited from Vanderbilt University. Participants were compensated with course credit.

Methods

All videos were presented as in Experiment 2. Half of the participants watched the twelve original videos shown in Experiment 1, the other half watched the same videos

with one major change – the videos ended immediately after the critical out of order clip (as defined in Experiment 3b) 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. After each of the twelve videos in both conditions, a black screen was displayed for 2000 ms, after which participants were shown a screen with two images labeled A and B. One image was taken directly from the critical out-of-order clip from the video they just watched. The other image was taken from footage not used in the final versions of the videos, but that could have reasonably been included, such as a shot from a different angle, or from immediately before or after the action took place. Participants were asked to respond whether they saw image A or image B at any point in the previous video. Immediately after completing this task, participants were prompted to respond whether or not they noticed an out of order action. All participants were given ten seconds to answer these questions, and the images were displayed throughout the entire response period.

Results

First, we investigated whether ending the videos immediately after the misordered action improved detection of the out of order event. On average, 70.83% (*SD* = 27.82) of the misorderings were detected in the abbreviated videos, while only 33.33% (*SD* = 19.24) were detected in the full-length videos. For videos that did not contain an out of order event, an average of 90.48% (*SD* = 16.27) of the full videos, and 87.50% (*SD* = 17.25) of the abbreviated videos generated correct responses. While there was a

significant effect of video length on accuracy in the misordered videos ($t(13) = 2.99$, $p = .01$), this was not the case in the normal ordered videos ($t(13) = -.34$, $p = .74$).

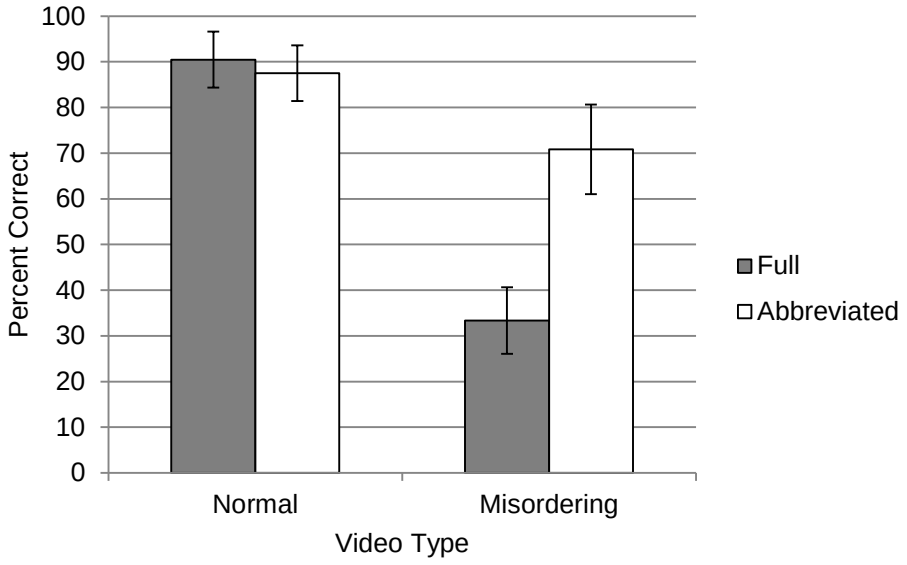

Figure 4. Effect of Video Type and Video Length on Correct Video Classification

Next, we examined the effect of video condition on memory for the target clip. In videos containing an out of order action, participants chose the correct image an average of 66.67% ($SD = 33.33$) of the time in full length videos, and 87.50% ($SD = 17.25$) of the time in abbreviated videos. In videos containing a correctly ordered sequence of events, participants chose the correct image an average of 76.19% ($SD = 25.20$) of the time in full length videos, and 91.67% ($SD = 15.43$) of the time in abbreviated videos. Whether a video contained a correct or misordered sequence did not significantly affect accuracy on the subsequent memory test ($F(1,13) = .57$, $p = .46$). Video length did affect recognition accuracy ($F(1,13) = 5.10$, $p = .04$), but it is important to remember that in the abbreviated versions of these videos, the target clip always immediately preceded the memory test.

This, however, was never the case for the full length videos. Finally, there was no significant video type by video length interaction effect ($F(1,13) = .09$, $p = .77$).

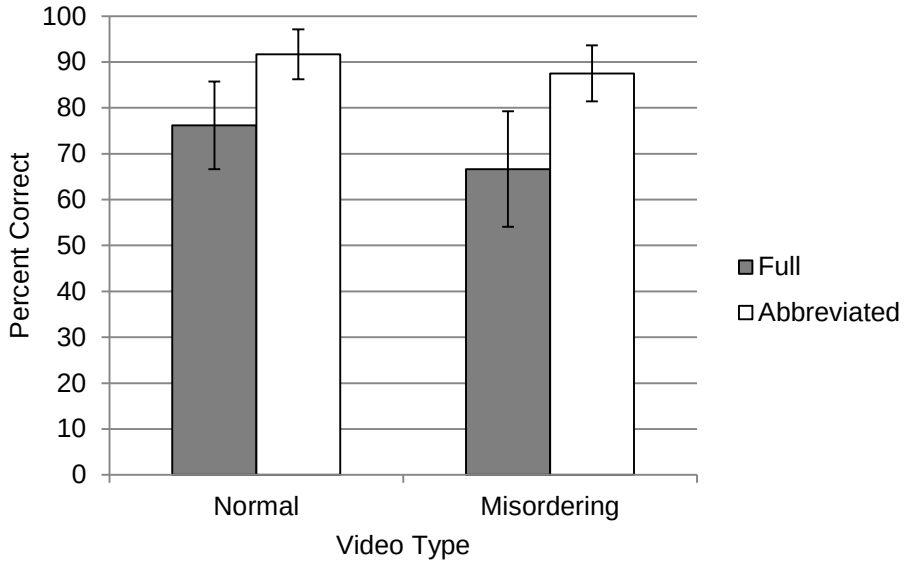

Figure 5. Effect of Video Type and Video Length on Memory Task Performance

Discussion

In this experiment, we found that stopping the videos immediately after the out-of-order clip significantly improved participants' ability to correctly detect the presence of misordered actions. This was not the case for correct rejections in videos that did not contain a misordering, however, this is possibly due to a ceiling effect, as performance in this condition was quite high overall. While video length also affected performance on the memory task, this was expected due to the shorter amount of time between the presentation of the to-be-remembered clip and the memory test in the abbreviated video condition.

The pattern of results observed in Experiment 4 supports the idea that the out-of-order action may indeed be in participants' working memory, yet they are unable to reliably use that information when perceiving an event. When participants were allowed to respond immediately after the out-of-order action, their ability to detect misorderings improved, perhaps because they were immediately given the opportunity to engage in additional processing, such as comparing the most recently seen action to previous actions. While participants did perform better on the memory test in the abbreviated video condition, we do not believe this can fully explain the increased performance on the main video classification task, assuming that event prediction and error detection is an automatic and on-line process. If it were truly an important and continuously occurring action, participants should have noticed the event misordering soon after it occurred, rather than waiting until the end of the video to reconstruct the event sequence and analyze their memory of the order of actions. Under this hypothesis, the length of the video sequence should have no effect on detection of misordered events, although it would affect one's memory for particular clips.

CHAPTER VI

GENERAL DISCUSSION

Across this series of experiments, we have come to a number of conclusions about people's ability to make automatic predictions continuously while perceiving an event and use these predictions to detect when an unexpected event occurs. First, our participants' performance on a misordering detection task decreased when they were asked to perform a secondary task. Second, their incidental performance on the task was lower than when they knew a misordering may be present. Third, their performance did not seem to be due to any difficulty in perceiving or understanding the stimuli. Finally, their performance improved when the event sequence stopped immediately after the unexpected event, allowing them to immediately engage in more effortful recall-based strategies, and removing any potential memory effects caused by the subsequent clips. Overall, while predictive processing is important in many areas of perception, it does not appear that moment-to-moment conceptual predictions in real-time event perception are used at the timescale represented by our stimuli.

While these findings have implications for the use of prediction in event perception in general, they have specific implications for Event Segmentation Theory (Zacks et al., 2007), which relies on an automatic prediction and error detection mechanism that operates at multiple timescales, simultaneously, to properly segment and ultimately understand events. For such a theory to be implemented, multiple requirements must be met. First, one's ability to segment an event depends on the perception of change

and violation of predictions. If nothing changes, then predictions should be easy to generate, and no event segments will be created. The scope of these predictions is left somewhat ambiguous. However, according to the model, this should include perceptual, such as motion, and conceptual, such as those related to an agent's intentions, predictions. While the on-line generation of perceptual predictions has been well established (e.g., representational momentum), the experiments described in this paper indicate that more conceptual predictions, such as predictions about an actor's future actions, may not be used as the model suggests.

Additionally, Event Segmentation Theory implies that event segmentation is an automatic part of perception that does not necessitate attention. This segmentation can occur simultaneously at multiple timescales, such as fine and coarse timescales. Therefore, event boundaries are perceived, even if they are not attended to. The results of Experiment 1 and 4 indicate that detection of unpredicted actions is not automatic, as it is greatly affected by a secondary task and improves when eliminating the presence of subsequent stimuli that may mask the memory of the out of order action. Real-time event segmentation would seem to require a completely automatic and effortless prediction, comparison, and model updating system, and the behavioral evidence does not necessarily support the hypothesis that these predictions are always automatic.

Finally, according to Event Segmentation Theory, violations of event models should be available to conscious awareness, as the models are what enable individuals to understand the events in their immediate environment. When the models are violated, understanding fails. If a perceiver is not aware that the event has changed, then event understanding becomes difficult, if not impossible, as they no longer have the correct

event schema in place and would be continuously generating incorrect predictions based on outdated beliefs. In the experiments reported here, participants found it very difficult to detect the event misordering, particularly in the incidental detection task of Experiment 3b. Interestingly, this experiment also showed some preliminary evidence for the presence of conceptual predictions that are not used. In making premature responses to the detection of the out-of-order clip itself, but not noticing the misordering, it seems as if occasionally participants expected the action to occur at a different time in the event sequence, but did not notice when reality failed to meet their expectations. In a way, it is as if there may be conceptual prediction that does not necessarily have the error-checking mechanism necessitated by models such as Event Segmentation Theory.

There is additional evidence in the literature to suggest that this conscious awareness of an order violation is not guaranteed. Raisig, Welke, Hagendorf, and van der Meer (2010) showed participants images representing three actions that make up an event (e.g., "order food" then "eat food" then "pay the bill"). These actions typically occur in a particular order. However, on some trials the triplets were shown in an incorrect order. This temporal violation induced an increased pupillary response in all participants, which was evidence of increased processing at the time the misordering was viewed. The authors argued that this response provided evidence that participants detected the temporal violation. However, not all participants were able to consistently report that a violation had taken place.

One potential extension of the experiments described in this paper would be a recreation of the incidental detection paradigm, with pupillary response as an on-line, non-explicit measure of detection. This could be used to determine the prevalence of

temporal violation detection that the participant is unable to later use in generating an explicit response. However, if this method is indeed measuring the error detection process that occurs in Zacks' model, as argued by the authors, then it too does not appear to consistently lead to the necessary automatic updating of awareness.

In conclusion, while both perceptual and conceptual predictions are important in a variety of day to day cognitive tasks, the experiments presented here indicate that online conceptual predictions are not important in typical event perception, as predicted by some models. If predictions are indeed generated, it seems unlikely that we are able to reliably use them to detect when an unexpected action occurs. These findings have implications for basic models of event perception, and suggest an underlying process that is not as closely tied to error checking the moment-to-moment sequential structure of events.

REFERENCES

Apperly, I.A., Riggs, K.J., Simpson, A., Chiavarino, C., Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, 17(10), 841-844.

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and Intentionality: Foundations of Social Cognitio*n (pp. 193-206). Cambridge, MA: MIT Press.

Freyd, J.J., & Finke, R.A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126-132.

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, 34(6), 1221-1235.

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *American Journal of Psychology*, 57(2), 243-259.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12(5), 436-450.

Newtson, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35(12), 847-862.

Premack, D., & Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, 4, 515-526.

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.

Vinson, N.G., & Reed, C.L. (2002). Sources of object-specific effects in representational momentum. *Visual Cognition*, 9(1/2), 41-65.

von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Head tracking and reaching for moving objects. *Cognition*, 76, 255–285.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28(6), 979-1008.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin*, *133*(2), 273-293.

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