

Running head: EVENT TEMPORAL RELATIONS

An Empirical and Computational Investigation of Perceiving and Remembering
Event Temporal Relations

Shulan Lu

Texas A & M University-Commerce

Derek Harter

Texas A&M University-Commerce

Arthur C. Graesser

The University of Memphis

Send Correspondence to:

Shulan Lu
Department of Psychology
Texas A&M University-Commerce
Commerce, TX 75429
Tel: (903) 468-8628
Fax: (903) 886-5510
Shulan_Lu@tamu-commerce.edu

Abstract

Events have beginnings, ends, and often overlap in time. A major question is how perceivers come to parse a stream of multimodal information into meaningful units and how different event boundaries may vary event processing. This work investigates the roles of these three types of event boundaries in constructing event temporal relations. Predictions were made based on how people would err according to the beginning state, end state, and overlap heuristic hypotheses. Participants viewed animated events which include all the logical possibilities of event temporal relations, and then made temporal relation judgments. The results showed that people make use of the overlap between events and take into account the ends and beginnings, but they weight ends more than beginnings. Neural network simulations showed a self-organized distinction when learning temporal relations between events with overlap versus those without.

Keywords: Event temporal relations; Event representations; Event boundaries; Knowledge representations; Recurrent neural network; End states; Beginning states; Overlap heuristic.

An Empirical and Computational Investigation of Perceiving and Remembering
Event Temporal Relations

In everyday life, we often experience and enact more than one event at a given time. Consider Tim and Pam at dinner. The waiter sets dishes on the table, and Pam moves a wine glass out of the way. Each of these events takes place over a time span with a particular duration. The event of *moving the wine glass* takes a few seconds, whereas *setting the dishes on the table* would typically take a minute or more. Events also have temporal relations with respect to each other. The event of *setting the dishes on the table* co-occurs to some extent with the event of *moving the wine glass*. Does Tim notice that the waiter situated the plate as soon as Pam moved the glass out of the way? Or does Tim only perceive that one event occurred after the other and infer the temporal relations based on the knowledge of event durations?

Studies have consistently shown that people's experiences of the world consist of segmenting the ongoing flow of perceptual information into discrete events instead of passively responding to stimuli (Zacks & Swallow, 2007). Event boundaries are of utmost importance in event perception, because they are the breakpoints where significant changes in both perceptual and conceptual representations take place. A major question is what role event boundaries have in representing event temporal properties. Based on the Newtonson (1973) task, Zacks, Tversky, and Iyre (2001) asked perceivers to segment everyday events under coarse versus fine grained encoding conditions. Consider a person *making a bed*. The results showed that the boundary of a fine grained event *straightening the sheet* aligned hierarchically and temporally with the boundary of the coarse grained event *putting on the sheet*. When segmentation data were aggregated across participants, the psychological event boundaries tend to cluster in a small interval for the majority of participants who marked the event juncture.

There is a natural inclination for people to assume that events (within some level of a hierarchy) tend to follow one another sequentially, with one event starting after the previous one ends. However, it is quite apparent when one takes a careful look at events in the real world that they do not transpire in such neat and tidy structures and sequences. There is considerable overlap and irregular timing among events, particularly when they are initiated by agents who attempt to execute plans in a complex physical and social world (Allen, 1984; Schank, 1999). Consider once again the example with Tim and Pam in the opening paragraph. Tim's perceptual experiences of *setting up the dish* and *moving the glass* are indexed by two separate agents (i.e., the waiter versus Pam). These two agents have separate plans that run along two separate tracks (i.e., events that occur in service of a particular agent's goal driven plan). Moreover, the events on one track may or may not be coordinated with the events on another track when there are multiple agents. It is easy to see how the relative timing of events ends up deviating from a hierarchically structured sequential flow of events.

The current paper describes three studies that assessed whether people tend to make event temporal relation errors that are biased toward any of the following three types of boundaries: the beginnings, ends, and boundaries where events either overlap or break apart from each other temporally (i.e., *overlap* versus *no overlap*). Systematic investigation of these errors is expected to shed light on how people mentally construct dynamic event representations. For example, if Tim performs perceptual segmentations on the two tracks simultaneously, then he would perceive this ongoing experience as one visual event. If Tim is limited to parsing one event track at a time, then he would perceive two separate tracks unfolding. These two alternatives assume that the mental representations of event temporal relations consist of two or perhaps all three types of event boundaries. The third alternative is that mental representations of event temporal relations are very

sketchy, for example, humans may only encode whether one event ends after another. These different possible mental representations of event temporal relations should be reflected in the error biases toward the beginnings, the ends, and the boundaries where events have interplay with each other.

Event Temporal Relations

In the fields of artificial intelligence and computational linguistics, Allen (1984, 1991) proposed a representation that contains seven relational primitives, with each primitive having two intervals over which two events occur. Fig 1 provides an illustration of Allen's seven event temporal relations. The relation between each pair of events is described by one of the seven predicates. The BEFORE relation means that one event is prior to another event and that the two events do not overlap in any way, whereas OVERLAP means that two events share part of the time course but have different beginnings and ends. It is important to note that the OVERLAP relation is not equivalent to *the overlap distinction* that we discuss throughout the paper. Following Allen, the OVERLAP relation is a precisely defined primitive relation that excludes the situation when two events have the same beginning or end points. In contrast, the overlap distinction refers to the ability to detect whether two events have any points in their time span that are coterminous; they may or may not have the same beginnings or ends. Other than OVERLAP, START, DURING, FINISH, and EQUAL relations have some degree of coextensive durations, whereas BEFORE and MEET have no temporal co-location.

Allen's representations have been used as basic temporal operators for automated planning and reasoning systems that make logical deductions about temporal relations. Studies in psychology provide some evidence that such interval-based instead of point-based representations are a natural way of relating events and drawing inferences about event temporal relations. For

example, people make relatively good estimations of event durations (Golding, Magliano, & Hemphill, 1992; Loftus, Schooler, Boone, & Kline, 1987), whereas people are not very good at estimating the points of time when events take place (Friedman, 1993; Golding et al., 1992; Linton, 1975).

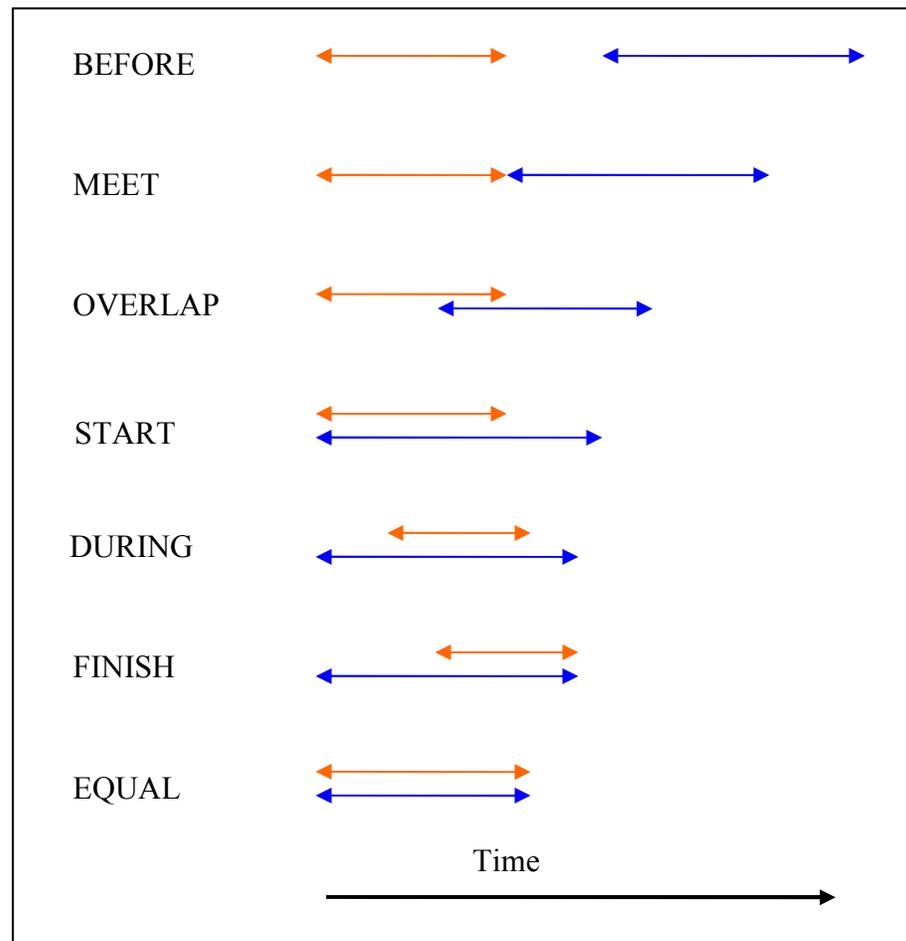


Fig 1.

Temporal Relations (Allen, 1991, pp.5).

Are events in everyday life amenable to Allen's representations? Some events have crisp, clear-cut temporal frames, which make it easy to define the temporal relations. *Chopping an onion* is an example, where the beginning, the end, and the temporal trajectory are not ambiguous.

However, many events invite problems, such as *throwing the baseball* (Davis, 1970; Thomson,

1971). Is the beginning best defined as the moment the pitcher grips the ball and winds up? Or is the beginning the moment the hand releases the ball? Aside from the inherent fuzzy temporal frame of events, the encoding and retrieval conditions of events are not always ideal, adding further sources of uncertainty to the processes (Chessman, 1985, 1988; Pearl, 1988). For example, people are frequently in the situation of perceiving multiple events, so they may not have enough cognitive and attentional resources to capture the beginnings and ends of all important events that occur during a dynamic situation.

People also tend to make inferences about temporal dynamics on the basis of causal relations and goal completions of events (Lichtenstein & Brewer, 1980). For example, infants often place events in an order that maps onto the logical inferences from causal constraints instead of the actual order in time (Bauer & Mandler, 1989). Adults perform in a similar manner as well. The event of *eating the meal* is perceived to occur after the event of *drinking a cocktail* because people impose simple causal or conventional structures on the events (Davidson, 1970 / 1996; Michotte, 1946 / 1963; Scholl & Nakayama, 2002; Wolff, 2003). Although conventional dining in the United States place cocktails prior to feasting on the main course, it would be entirely possible for the diner to drink the cocktail intermittently while working on the main course. If so, the two events would have an OVERLAP relation instead of a BEFORE relation.

In summary, the psychological frame of an event may deviate from its physical frame. The relative intervals of different event boundaries are particularly important for judging event temporal relations. For example, humans could encode and store the fact that one event ends later than another, and use this information to infer the event temporal relation. The metrical property of these relative temporal intervals could be categorical (e.g., earlier, later, apart, overlap). This is consistent with the principle that representations need not be restricted to absolute judgments of

quantitative details (Garner, 1962; Miller, 1956). In object recognition, for example, Biederman (1987) reported that the quantitative processing of the length of an object is slow and prone to errors compared with identifying the object itself. Below we describe three hypotheses that investigated how the relative intervals of the beginnings, the ends, and the interplay between events over multiple tracks contribute to the errors of perceiving event temporal relations.

The Overlap Heuristic Hypothesis

Previous studies reported that contextual information instead of specific time tags could account for the encoding and retrieval of the temporal order in which words are presented (see Friedman for a review, 1993). However, the contextual overlap account does not address how people segment continuous streams of multimodal information into meaningful events and how they use the event boundaries to process temporal relations. Reynolds, Zacks, and Braver (2007) recorded the motions of an actor enacting 13 routine actions and used these to form a sequence of actions. The sequence was then converted into matrices that consisted of the 18 position points of the actor's body in a 3-dimensional space. The neural network received a stream of such event inputs and predicted what the next input was. The results showed that significant influxes of perceptual prediction errors give rise to the perception of event boundaries. For BEFORE and MEET events, there is at least one moment that separates two events (i.e., no overlap), thus there should be significant spikes in prediction errors between these two relation categories. Similarly, there should be significant increases in prediction errors when a new track of activity begins or ends in the midst of a separate track of activity (i.e., when there is at least some coterminous overlap).

The overlap hypothesis predicts that humans are likely to confuse event temporal relations within the family of overlapping relations (OVERLAP, START, DURING, FINISH, and

EQUAL) and likewise humans are also likely to confuse event temporal relations within the family of non-overlapping relations (BEFORE and MEET); however, humans are not as likely to confuse event temporal relations across the two categories. There is some evidence supporting this prediction. Gennari (2004) reported that comprehenders' reading times for two-clause sentences are influenced by whether the events and states in the clauses overlap in time. However, neither the size of the overlap interval nor the size of the temporal gap between the events introduced additional effects on reading time. The values of 1 in Table 1 signify that errors are theoretically acceptable according to the overlap heuristic hypothesis, whereas the values of 0 stipulate that errors are unlikely according to the hypothesis. If the presented stimuli are BEFORE events as indicated in Row 1, then BEFORE events are very likely to be mistaken as MEET as indicated by the value of 1. But BEFORE events are not likely to be mistaken as any of the overlapping relations when an error is made, as indicated by the value of 0 in Table 1. No predictions regarding the correct judgments along the diagonals can be made on any of the hypothesis alone.

Table 1. *Overlap Heuristic Prediction*

| Stimulus Events | Human Judgment | | | | | | |
|-----------------|----------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | . | 1 | 0 | 0 | 0 | 0 | 0 |
| MEET | 1 | . | 0 | 0 | 0 | 0 | 0 |
| OVERLAP | 0 | 0 | . | 1 | 1 | 1 | 1 |
| START | 0 | 0 | 1 | . | 1 | 1 | 1 |
| DURING | 0 | 0 | 1 | 1 | . | 1 | 1 |
| FINISH | 0 | 0 | 1 | 1 | 1 | . | 1 |
| EQUAL | 0 | 0 | 1 | 1 | 1 | 1 | . |

The End State Bias Hypothesis

The attentional focus on the culmination of goals and plans has been a major theoretical approach in accounting for the mental organization of events (Lichtenstein & Brewer, 1980; Miller, Galanter, & Pribram, 1960; Mourelatos, 1978; Newell & Simon, 1972; Schank & Abelson, 1977). In this framework, everyday activities are comprehended in goal-directed partonomic hierarchies. Lichtenstein and Brewer (1980) suggested that people frequently reconstruct the temporal order between events based on plans and goals.

If the achievement of a goal is an important way of marking event junctures, then people are more likely to attend to the end states of events rather than the beginnings. There is some evidence for preferences to the end points of spatial motion events (Regier & Zheng, 2003). Participants were presented with two motion events at a time and were asked to make judgments whether or not the two presented events were the same. Participants made fewer errors when the events emphasized the event endings (e.g., *putting a lid on a container*) than when the events emphasized the beginnings (e.g., *taking a lid off a container*).

The end state bias hypothesis assumes that humans are more likely to represent temporal relations on the basis of the relative interval between the end states. Two events either end at different points in time or end at the same point in time. When considering Fig 1, BEFORE, MEET, OVERLAP, START, and DURING have relative intervals between the end states that are greater than zero, whereas FINISH and EQUAL have zero relative intervals between the end states. The end state bias hypothesis stipulates that BEFORE, MEET, OVERLAP, START, and DURING are likely to be confused with each other, whereas FINISH and EQUAL are likely to be confused with each other; however, these two clusters of primitive relations are less likely to be confused with each other.

Table 2 shows the predictions based on the end state bias. For example, the OVERLAP events in Row 3 may be mistaken as START and DURING which are indicated by the values of 1 in the matrix, whereas OVERLAP events would not be mistaken as FINISH and EQUAL as indicated by the values of 0. Considering the end state hypothesis exclusively, BEFORE and MEET could be mutually mistaken as OVERLAP (or START, DURING), since their end states precede one another. Clearly, Table 2 has a slight difference from the pure end state bias hypothesis we just described.

The rationale for Table 2 is that people may rule out or consider the overlapping relations as a possibility before they weight the end and the beginning states. Gennari (2004), for example, reported that readers establish whether events overlap or not immediately during comprehension. If we allow the overlap also to be a factor, along with the end state, then the OVERLAP relation would not be mistaken as BEFORE and MEET. Table 2 reflects this modified end state bias hypothesis, taking into account some allowance for the perception of overlap to influence the event temporal judgments.

Table 2. *End State Bias Prediction*

| Stimulus Events | Human Judgment | | | | | | |
|-----------------|----------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | . | 1 | 0 | 0 | 0 | 0 | 0 |
| MEET | 1 | . | 0 | 0 | 0 | 0 | 0 |
| OVERLAP | 0 | 0 | . | 1 | 1 | 0 | 0 |
| START | 0 | 0 | 1 | . | 1 | 0 | 0 |
| DURING | 0 | 0 | 1 | 1 | . | 0 | 0 |
| FINISH | 0 | 0 | 0 | 0 | 0 | . | 1 |
| EQUAL | 0 | 0 | 0 | 0 | 0 | 1 | . |

The Beginning State Hypothesis

In everyday life, it is not hard to imagine how we represent the beginning states of events somewhat differently from the actual beginning points in time. For example, the two parallel sessions of a conference may not start at the same point in time, but people tend to perceive them as beginning simultaneously. In philosophy, Hacker (1982 / 1996) noted the beginnings have differential status in the course of an event compared with the endings. However, compared with the emphasis on the end state in action, memory, planning, and language, there has not been much attention paid to the beginning states of events in psychology. Recently, Lakusta (2006) reported no significant bias toward the end states of intentional events, for example, an actor walks toward a goal object (e.g., a chair) but gazes at the beginning (source) object (e.g., a post). Reynolds et al. (2007) reported that the neural network made significantly more prediction errors at the beginnings of new events instead of the ends of the existing events. Subsequently they proposed that the event breakpoints are the beginnings of new events instead of the endings of old events.

Opposite to the end state bias, the beginning state hypothesis assumes that humans apprehend beginning states and make errors due to the confusions of the relative intervals between beginning points. BEFORE, MEET, OVERLAP, DURING have different beginning points in time and therefore are likely to be confused among each other, whereas START and EQUAL have the same beginnings and are likely to be confused. Table 3 shows the predictions made based on the beginning state hypothesis. Consistent with Table 2, Table 3 incorporates the overlap heuristic hypothesis in that BEFORE and MEET will not be mistaken as any overlapping relations.

Table 3. *Beginning State Prediction*

| Stimulus Events | Human Judgment | | | | | | |
|-----------------|----------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | . | 1 | 0 | 0 | 0 | 0 | 0 |
| MEET | 1 | . | 0 | 0 | 0 | 0 | 0 |
| OVERLAP | 0 | 0 | . | 0 | 1 | 1 | 0 |
| START | 0 | 0 | 0 | . | 0 | 0 | 1 |
| DURING | 0 | 0 | 1 | 0 | . | 1 | 0 |
| FINISH | 0 | 0 | 1 | 0 | 1 | . | 0 |
| EQUAL | 0 | 0 | 0 | 1 | 0 | 0 | . |

Overview of the Study

This work aims to understand how humans construct event temporal relations and investigates the relative weights of each of the three previously discussed hypotheses in temporal relation judgment errors. The sections below briefly describe the stimuli, the tasks, and the analysis that were used.

Zacks and Tversky (2001) suggested that events viewed at the time scale of a few seconds tend to be perceived as simple physical changes and that events viewed at the time scale of 10 to 30 seconds tend to be perceived as more intentional. Taking this into account, we started with an animated event of fish swimming that lasted 10 seconds or shorter. We combined two of these animated events to form one of Allen's seven temporal relations between them. The total duration of the two events lasted 25 seconds or shorter. For the animations in Experiment 1, as much as possible we (a) kept the amount of overlap between events approximately the same across the overlapping relations, and (b) maintained the amount of overlap shorter than the duration of either

of the two individual events. For the animations in Experiment 2, there were some variations in the amount of overlap as needed by the animation setups, ranging from 3 to 10 seconds of overlap in time. The weights of each type of confusion in the confusion matrix may vary, if the ratio between the event durations and the amount of overlap vary in a wide spectrum. To simplify the tests of theoretical predictions at this point, the confusion weight between OVERLAP and MEET was set at zero. However, it is likely that OVERLAP will be confused as MEET when the amount of overlap between two events is very short, such as 1 second. The animations can be found online at <http://faculty.tamu-commerce.edu/dharter/shulan/animations/shulan-animations.html>.

For the psychological task, participants first viewed the animations of fish swimming. They then received Allen's arrow diagram, and selected one relation that best depicted the animated events they just viewed. It is of course conceivable that the use of lines and arrows would introduce some emphasis in spatial processing. However, time is a domain that appears to be grounded in spatial visual representations (Boroditsky & Ramscar, 2002; Lakoff & Johnson, 1980). An alternative measure would be a language based task, which in turn may introduce a language bias. For example, the word *simultaneously* may depress the differentiations between various overlapping relations. To gauge whether Allen's diagram will introduce significant extra spatial processing, we performed a pilot study in which participants received the descriptions of fish swimming events and asked them to draw the event temporal relations without introducing Allen's relations ahead of time. Participants did not report any difficulty in doing so, suggesting that spatial representations are natural for representing temporal relations.

Given each type of stimuli, we could obtain the proportions of confusions among each of the seven relations. First, the proportions of each type of confusion errors were used as the criterion variable in regression, whereas the predictions made by each of the hypotheses were used

as predictors. We performed (a) a stepwise regression analysis using the predictions made by the three hypotheses, and (b) a separate hierarchical regression analysis using the predictions made by the beginning and end states when taking into account that people may establish whether events overlap. Second, we computed the similarity between each type of confusion and performed multidimensional scaling.

In addition to collecting data on human judgments of event temporal relations, we investigated whether a neural network model could account for the human judgments. In the family of artificial neural networks, patterns of temporal properties can be modeled by recurrent connections that save and propagate past states of a network to the current context (Botvinick & Plaut, 2004; Elman, 1990; Jordan, 1988). The current study used a simple recurrent neural network (RNN) and varied the overlap interval in a wider spectrum than the stimulus events used in Experiments 1-2.

Experiment 1: Perceiving and Judging Simple Events

Experiment 1 investigated the three hypotheses when two simple events were not related by virtue of goals or causal constraints. The animations consisted of two individual fish that swam in a simple environment: the fish swim or float independently. Participants viewed animations of the fish swimming events and then made judgments about the event temporal relations. The overall error rates in Experiment 1a were low, so we conducted a follow up experiment that produced higher error rates. Experiment 1b used the same materials and procedure as Experiment 1a, but resampled the animations to be 2.5 times faster.

Method

Participants. Fifty-one introductory psychology students at the University of Memphis participated in Experiment 1a for course credit, whereas an additional forty-five students

participated in Experiment 1b for course credit.

Materials. Forty-two animations were developed using 3D Studio Max, release 5. In each animation, two fish of different color and size start out from two separate corners and swim above a grid within a 3D environment of a fish tank (Appendix 1). For a given animation, the two fish swim at different speeds but over the same distances, or two fish swim at the same speeds but over different distances. The grid has four sides, so there are four orientations in which the fish swim parallel to each other along the net, and two orientations in which the fish swim intersecting to each other.

The animation quality was near photorealistic: All colors were prepared with texture mappings. A snapshot of an example animation is shown in Appendix 1. Experiment 1a presented the animations at the rate of 30 frames per second, whereas Experiment 1b presented the animations at the rate of 75 frames per second. There was a correct classification of event temporal relation as defined by the frames of each event in each animation.

Procedure. Each participant was seated in front of a Pentium computer, which used MediaLab to display the experiment materials (Jarvis, 2000). Participants were told that they would be asked to make judgments concerning how fish swimming events were related in time. Participants were shown a diagram similar to Fig 1, except that the word labels (e.g., BEFORE) were absent. Participants were asked to tell the experimenter the meaning of the seven temporal relations depicted in the diagram. The experimenter provided feedback to each participant about the understanding of each relation. The experimenter did not launch the animations until participants understood all seven relations.

Participants viewed each animation only once. Following the disappearance of the animation, participants received a diagram with Allen's seven relations. Participants chose one

relation out of the seven that best captured how the two animated events were related in time.

Participants made choices by clicking a number that was next to the temporal relation. There were 20 sets of orders in which the seven temporal relations were listed in a diagram. Each participant received one of the 20 orders and used the same order throughout the experiment. Each participant viewed 42 animations in a random order. The randomization was completed by MediaLab 2000.

Error Analysis. Two sets of error analyses computed which temporal relations people tend to confuse. The first set of analysis directly calculated the proportions of errors on each of 7 temporal relations. The second set of analysis calculated the similarity among the 7 temporal relations using an entropy measure, and then a similarity matrix was constructed for multidimensional scaling.

Entropy measures the amount of uncertainty about a given outcome (Shannon, 1948). If people often confuse two particular event temporal relations, then the uncertainty between the two relations is greater than other relations and the two relations are more similar to each other than to other relations. For each stimulus item, participants' judgments formed a distribution containing seven values (the proportions of judgments in each row). There are seven types of stimuli, thus they form a 7 by 7 response matrix.

For a given type of stimulus item, the following formula computes how likely one relation is confused with another one.

$$E_i = -\frac{\sum_{i=1}^N p_i \ln p_i}{\ln N} \quad (\text{Entropy Formula})$$

The p_i refers to the proportion of times a given choice is selected out of the N ($N = 7$) possible items. The entropy provides an index of how similar any two given relations (e.g., BEFORE versus MEET) appeared to participants. A similarity matrix can thus be constructed, and then

analyzed using multidimensional scaling (MDS). The program was implemented in SYSTAT version 10 with Kruskal S-STRESS scaling method. The results exhibit the organization of temporal relation errors.

Experiment 1a Results

Table 4 presents the proportion scores of the seven alternative judgment tasks in Experiment 1a. For example, the BEFORE row indicates that participants correctly identified BEFORE events 89.5% (.895) of the time. Participants confused BEFORE events with MEET events 4.9% (.049) of the time, and so on. The overall error rates of the seven relations were the following: BEFORE (.10), MEET (.22), OVERLAP (.37), START (.16), DURING (.36), FINISH (.22), and EQUAL (.05). Our analyses assessed whether these empirical proportions could be predicted by the theoretical matrices presented in Tables 1, 2, and 3. The theoretical matrices were normalized so that the sum of the error cells within a row would be 1.0. That is, each error cell in a given row was first added up, and then divided by the sum of the row. The normalized matrices were used to perform the analyses throughout this paper.

Table 4. *Proportion of Temporal Judgments Experiment 1a*

| Events in the World | Human Judgment | | | | | | |
|---------------------|----------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | .895 | .049 | .007 | .013 | .020 | .010 | .007 |
| MEET | .046 | .778 | .092 | .010 | .033 | .043 | .000 |
| OVERLAP | .013 | .010 | .628 | .072 | .163 | .095 | .020 |
| START | .003 | .003 | .059 | .843 | .029 | .012 | .043 |
| DURING | .013 | .016 | .105 | .134 | .637 | .069 | .026 |
| FINISH | .013 | .003 | .056 | .059 | .039 | .778 | .052 |
| EQUAL | .013 | .003 | .007 | .013 | .003 | .010 | .951 |

Two multiple regression analysis were performed on the normalized matrix of proportions. First, a stepwise regression was performed to investigate which one of the three theoretical predictors made the most contributions to the errors. The predictors included: (a) overlap hypothesis as in Table 1, and (b) the end state and the beginning state hypothesis that did not take into account of the overlap hypothesis. The normalized proportions of errors in Table 4 served as the criterion variable. This analysis investigated the extent to which people segment two tracks of events and use the segmentation boundaries in constructing event temporal relations. The data could not ascertain the exact locations of the event breakpoints. However, if there is no trace of such event processing, then either the relative beginning or ends should come out as the primary predictor of the regression. The overlap hypothesis accounted for 27% of the variance, $R^2 = .27$, $F(1, 40) = 15.09$, $p < .0001$. The end and beginning state hypotheses added only 6% of the variance.

Second, keeping the same criterion variable, a separate hierarchical regression was performed by using the values from the theoretical matrices for the end state and the beginning state bias hypotheses as in Tables 2 and 3. This analysis investigated the role which the relative beginnings versus ends have in constructing event temporal relations when the boundaries that signal the dynamics of two events come into play. The two predictors were not significantly correlated, $r = .21$. The two predictors in this hierarchical regression accounted for 50% of the variance, $R^2 = .50$, $F(2, 39) = 19.18$, $p < .0001$. The end state ($\beta = .49$, $p < .0001$) showed some trend of having a greater effect on the errors than the beginning state bias ($\beta = .41$, $p < .001$).

To further examine the roles of the three types of event boundaries, we performed additional hierarchical regression analysis, using the overlap heuristic hypothesis in Table 1 and the end state bias hypothesis in Table 2. The results showed that there was an 11% increase in variance accounted for by the end state bias hypothesis over and above the overlap heuristic

hypothesis, whereas there was only a 6% increase in variance accounted for by the beginning state hypothesis over and above the overlap heuristic hypothesis. Such pattern of increases indicated that not only the overlap but also the end state guides the construction of event temporal relations.

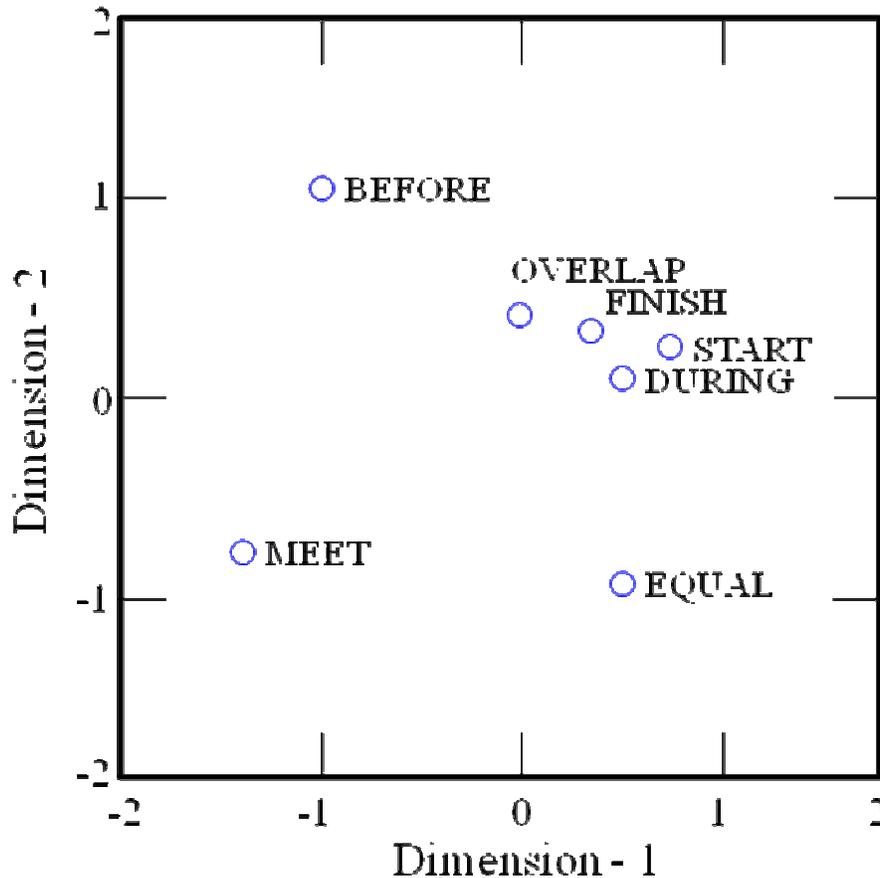


Fig 2.

MDS solution for Experiment 1.

Fig 2 shows the MDS solutions of how similar event temporal relations appeared to participants. Based on the entropy values, a similarity matrix was constructed. Then the similarity matrix was fit by a 2-dimensional MDS solution, with a low stress value (.03). This MDS analysis addressed the question along which dimension people sorted event temporal relations. If simultaneous segmentation does not occur at all, for example, the beginning state was the sole

driving force, there should be the following clusters: (BEFORE, MEET, OVERLAP, FINISH, DURING), versus (START, EQUAL). The seven temporal relations fell into disjoint groups: BEFORE; MEET; (OVERLAP, FINISH, START, DURING), and EQUAL.

Analysis of Experiments 1a and 1b

Compared with Experiment 1a, the animated events in Experiment 1b were presented 2.5 times faster. Although the error rates in Experiment 1b increased by 12%, the proportions of corrected judgments and confusions in Experiment 1b held the same pattern as Experiment 1a, $r(49) = .99, p < .001$. The stepwise regression analysis and MDS results also held the same pattern as Experiment 1a. In Experiment 1a, the end state did not have a huge advantage over the beginning state in accounting for errors. We were interested in seeing whether the increase in task difficulty and error rates would make any difference. In Experiment 1b, the hierarchical regression showed that the end state bias had greater effects on the errors made ($\beta = .50, p < .001$) than the beginning state ($\beta = .31, p < .05$).

In Experiments 1a and 1b, events that had the same beginnings were the first stimuli that were launched and therefore may have primacy effects, whereas events that had the same endings were the last part of the stimuli and may have recency effects. It is difficult to ascertain whether the performance on those relations is due to primacy or recency effects or due to some cognitive bias toward a particular type of event boundaries. If, for example, there is a cognitive bias toward the end states instead of recency effects, then people would always try to capture the ends whether events end at the same time or not. This would predict that the performance on events that had the same ends versus different ends would be lower as we go from Experiment 1a to Experiment 1b. If, for example, there is a recency effect, then people would perform worse on events that had different ends.

We performed an analysis of variance of Experiments 1a and 1b together on the proportions of correct judgments, using an end state (different end points versus same end points) by presentation speed (normal versus faster) design. There was no significant interaction, $F(1, 94) = .30$. The hit rates on the relations that had the same ends dropped by .12 (.86 in Experiment 1a versus .74 in Experiment 1b), whereas the hit rates on the relations that had the different ends dropped by .11 (.76 in Experiment 1a versus .65 in Experiment 1b). These results suggested that there is some cognitive bias toward the end states and this is not altered by processing load.

Following the same design, an analysis was performed regarding the beginning state. There was a significant interaction, $F(1, 94) = 5.37, p < .05$. For relations that had the same beginning states, the hit rates dropped by .05 (.89 in Experiment 1a versus .84 in Experiment 1b), whereas for relations that had different beginning states, the hit rates dropped by .13 (.74 in Experiment 1a versus .61 in Experiment 1b). These results suggested that the primacy effect rather than cognitive bias accounts for the performance on beginnings.

An analysis using the same design was performed regarding the overlap versus no overlap. If event temporal relations consist of only one type of event boundary, for example, the beginning states, people can easily capture that one event occurs after another for non-overlapping events. However, if people try to encode the event boundary where events break apart from each other, then the increase in task difficulty should significantly impact the hit rates for non-overlapping relations. In contrast, if the overlapping event boundaries were part of the event representations, then people should always try to attend to the overlapping event boundaries, and thus the hit rates for coterminous events should not go significantly lower when events are presented faster. There was indeed a significant interaction, $F(1, 94) = 14.75, p < .0001$. For non-overlapping relations, the hit rates dropped by .20 (.83 in Experiment 1a versus .63 Experiment 1b), whereas for

overlapping relations, the hit rates dropped by .08 (.77 in Experiment 1a versus .69 Experiment 1b). These results suggested that event temporal relations contain more than one type of event boundaries. People encoded how events break apart from each other and how events overlap with each other.

Discussion

Experiment 1 demonstrated that Allen's temporal representation has some degree of psychological plausibility because participants made temporal relation judgments reasonably well. The results ruled out one possibility, namely that humans judge the temporal relations with the relative beginnings and ends only, and infer whether events overlap via the relative beginnings and ends. Instead, humans tend to construct temporal relations from event boundaries in a fashion that first computes whether events overlap or break apart from each other and subsequently computes the status of the end states. It is not clear to what extent humans encode and remember when events begin.

We performed an analysis to further ascertain the extent to which the task itself introduced complex spatial processing that could significantly affect the results. For each temporal relation, the animated events were either spatially parallel or intersecting. If the spatial processing introduced heavy distortions into the results, participants should perform worse on the animated events that were spatially intersecting. The overall proportions of correct judgments on animated events that were spatially intersecting (.78 in Experiment 1a) were not lower than animated events that were spatially parallel (.79 in Experiment 1a). The results were the same in Experiment 1b (.68 and .66). This data did not suggest that the dependent variable introduced significant adverse spatial biases.

Participants tended to distinguish event temporal relations along the lines of having

overlap or not. However, the overlap heuristic hypothesis did not tell the whole story. The results suggested a trend in which the end state played a slightly bigger role than the beginning state in explaining the errors. Is there some inherent bias in the stimuli that might explain such results? We classified the target test events in Experiment 1 into three types: (a) those that have end states far from the animation frame, (b) those that have end states near the frame, and (c) those that have end states in the middle of the frame. In Experiment 1a, for example, the proportions of correct judgments on each of the above three types of events were .79, .80, and .78. This data did not suggest a particular bias towards the beginning, end, or the middle portion of the animation.

The proportions of correct judgments on EQUAL were significantly higher than the rest of the overlapping relations. One may wonder if any other factors could account for the performance. The simple EQUAL events in Experiment 1 did not have events before and after, so they may have benefited from both primacy and recency effects. If that were the case, having EQUAL events embedded in a longer stream of events should eliminate the effects. The results in Experiment 2 will demonstrate this.

The results raised a question of whether attentional resources could accommodate simultaneous segmentation. Studies on attentional blink showed that it takes half a second for people to adapt to the second stimulus. However, Olivers (2007) reviewed a number of studies showing that the attention gate starts to close if the second stimulus is a distractor, whereas the attention gate remains open when the second stimulus is relevant to the task. Yamada and Kawahara (2007) demonstrated that perceivers could simultaneously monitor up to four objects at two noncontiguous locations rather than rapid shifting of attention. These studies in attention suggested that people could mark down the overlapping event boundaries. For example, people performed well for event temporal relations that had the same beginnings (START). This result

was consistent with the finding that the processing of the second stimulus is not affected if the stimulus launches within the 100 milliseconds after the first stimulus.

In Experiment 1, each animation had two events and each event had a crisp beginning and end. A question arises as to whether the results could generalize to situations in which multiple events unfold and events have somewhat fuzzier beginnings and ends. If indeed people attend to the event boundaries, then the results exhibited in simple fish swimming events should scale up to situations in which multiple events are linked to each other by causal relevance.

Experiment 2: Perceiving and Judging Events Embedded in a Schema

Studies have consistently shown that an overarching knowledge structure has an impact on event representations (Abott et al., 1985; Anderson & Pichert, 1978; Graesser, Kassler, Kreuz, & McLain-Allen, 1998; Nelson, 1986; Rumelhart & Norman, 1985; Schank & Abelson, 1977). Bower et al. (1979), for example, is a classic study on the comprehension and memory of events in everyday activities. The results showed that people tend to recall events in an order that corresponds to the underlying event sequence, even though the events were mentioned out of order in the original text. Such findings suggest that perceivers may rely more on conventional knowledge structures to propagate inferences and build mental representations. The critical question is whether the cognitive system tunes to the more conceptually driven part of the event temporal relations. An active maintenance of the knowledge representations is required for perceivers to attend to the goals states while facing the competition of irrelevant information. It is conceivable that perceivers have a perceptual bias toward the end states, which are frequently aligned with goals. In particular, when events are more complex, the bias toward end states may become more pronounced.

Experiment 2 investigated the extent to which the overlap, the end state, and the beginning

state hypotheses control the errors in temporal relations when events are embedded in a higher order structure as opposed to the direct level of focus as in Experiment 1. Experiment 2 embedded test events in a *schema*. Schema is a generic term for a conventional, higher-order, knowledge structure. The story schemata in Experiment 2 were derived from the social events in a study conducted by Morris and Peng (1994), which examined the causal attributions to physical versus social events. For example, a big fish scared away a group of small fish near a plant. The high order structure involves a fish scaring away other fish, whereas the fine-grained level specifies the particular events of motion. The stimuli set systematically varied the temporal relations between one particular fish swimming event and one particular member of a fish group swimming event. As in previous experiments, participants made judgments of temporal relations after they viewed each animation.

Method

Participants. Forty introductory psychology undergraduates at the University of Memphis participated for course credit.

Materials. Three sets of animations were made using 3D Studio Max. There were 7 animations in each set, which had two events that illustrate the temporal relation in Allen's representation. An example animation is shown in Appendix 2.

(1) The chasing schema set illustrated the conflict between a group of fish, headed by an orange fish with green spots, and a blue-patched fish which was already at the target plant. At the sight of the orange fish, the blue-patched fish swam back to its habitat. The animations varied the timing of the two events: One event referred to the orange fish darting forward with the group, whereas a second event referred to the blue fish leaving the target and then swimming to the plant on the far left. Each animation for this stimulus set contained the above two events, and depicted

one of Allen's temporal relations, resulting in seven animations in this set. The other two sets of animations followed this design.

(2) The joining schema set illustrated the rejoining of a member to its group. One group of fish, led by an orange fish with green spots, headed toward the targeted plant, whereas one member lagged behind the group. The animations varied the timing of the two events: One event referred to the orange fish leaving the target plant and joining the group, whereas the second event referred to the blue-patched fish leaving the target plant and rejoining the group.

(3) The dispersing schema set illustrated that one group of fish, headed by an orange fish with green spots, swam to a target plant, and scattered in the presence of a big blue fish. The blue fish slowly emerged out of its habitat, left its habitat, and swam toward the target plant. As a result, the group of fish started to head back to their habitat. The group members did not leave the plant at the same time. The orange fish did not leave until every other fish leaves. The animations varied the timing of the two events: One event referred to the orange fish heading back toward the group habitat, whereas a second event referred to the blue fish swimming from its habitat to the target plant.

Procedure. A pilot study was conducted to ensure the test events match with participants natural event perception. First, three participants simply described the events while watching the animations. The results showed that 89% of the time both test events were described. Second, an additional 10 participants described each event whenever they thought there was a meaningful event in the animation, and later they judged the temporal relations of two target events.

Participants reported both events 93% of the time.

The procedure in Experiment 2 was essentially the same as the one used in Experiment 1, except that participants were told that each animation had a story before the experiment began.

More specifically, participants were verbally told the following: in one story, some fish try to chase away another fish; in another story, a fish scares away a group of fish, or alternatively a fish tries to rejoin its group. During the experiment, participants were not provided any written text about each animation.

The animations were presented to participants one at a time in a random order. After viewing each animation, participants received the event descriptions regarding the two target events on a separate computer screen. Participants then clicked a number next to the temporal relation that best captured the test events. Below are example event descriptions.

The big fish event: the big fish swims to the plant in the center.

The fish with green dots event: the fish with green dots turns its head and swims to the plant on the right.

Results

Table 5 presents the proportion scores for the seven alternative relation judgment task in Experiment 2. Two separate multiple regression analysis were performed as in Experiment 1. The stepwise regression did not keep the beginning state hypothesis in the regression equation. The overlap heuristic hypothesis and the end state hypothesis together accounted for 42% of the variance of the error proportion scores in Table 5, $R^2 = .42$, $F(2, 41) = 14.16$, $p < .0001$. The overlap heuristic hypothesis ($\beta = .53$, $p < .0001$) had a significantly greater effect on the errors as compared with the end state hypothesis ($\beta = .32$, $p < .05$).

As in Experiment 1, the overlap heuristic hypothesis was the major predictor, but would the beginning state still play a significant role once considering that participants may have first established whether events overlap? The hierarchical regression showed that the end state bias and the beginning state hypotheses, as described in Tables 2 and 3, accounted for 34% of the variance,

$R^2 = .34$, $F(2, 39) = 10.21$, $p < .0001$. The beginning state ($\beta = .26$, *ns*) was not a significant predictor.

Table 5. *Proportion of Temporal Judgments in Experiment 2*

| Events in the World | Human Judgment | | | | | | |
|---------------------|----------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | .258 | .300 | .208 | .100 | .050 | .067 | .017 |
| MEET | .158 | .292 | .267 | .133 | .050 | .083 | .017 |
| OVERLAP | .050 | .083 | .417 | .100 | .117 | .142 | .092 |
| START | .025 | .058 | .225 | .308 | .175 | .117 | .092 |
| DURING | .075 | .050 | .158 | .175 | .125 | .233 | .183 |
| FINISH | .083 | .058 | .167 | .067 | .117 | .350 | .158 |
| EQUAL | .058 | .033 | .142 | .142 | .092 | .267 | .267 |

Fig 3 shows the MDS solutions. The similarity matrix was fit by a 2-dimensional MDS solution, with a low stress value (.14). BEFORE and MEET are apart from other temporal relations, but MEET appears to be closer to OVERLAP than in Experiment 1. OVERLAP, (START, FINISH, DURING), and EQUAL are closer to each other than to BEFORE and MEET. START, FINISH, and DURING are closer to each other than to OVERLAP and EQUAL. The results are compatible with the conclusion that event temporal relations clustered along the overlap hypothesis, but also indicted how the end states played a role (i.e., the closer distance between FINISH and EQUAL compared with Experiment 1).

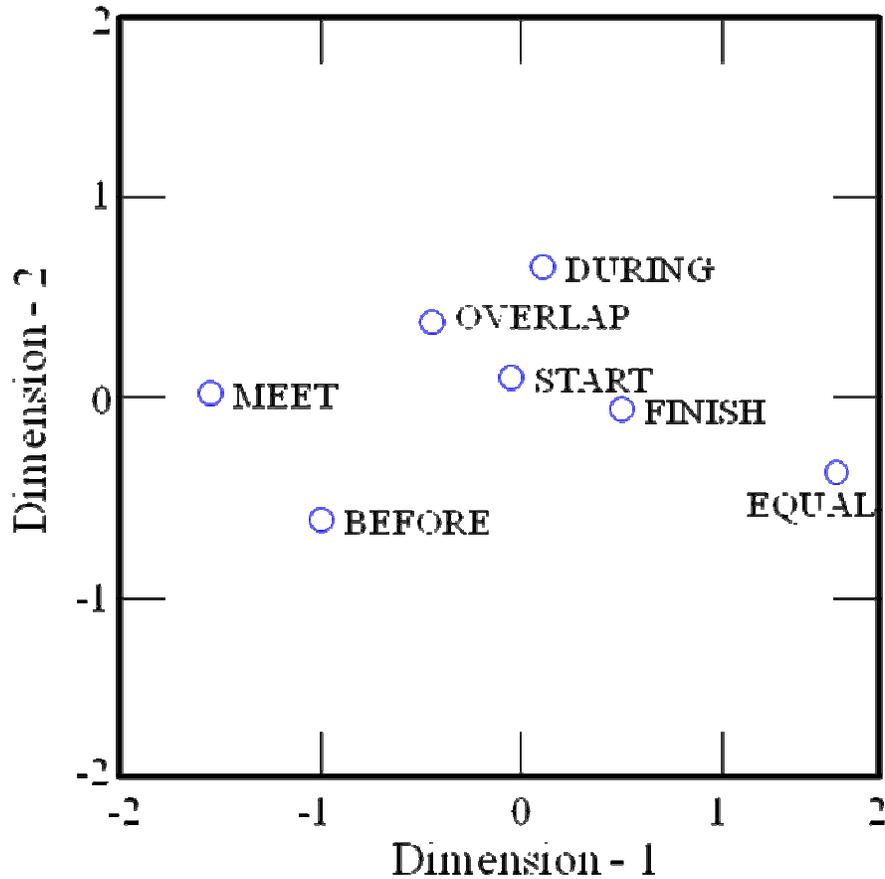


Fig 3.

MDS solution for Experiment 2.

Discussion

In Experiment 2, participants viewed animations of fish swimming events that were embedded in streams of events organized by a schema. The results showed that the temporal relations were not simply deduced based on the causal relations. For example, a fish typically does not leave a habitat until a threat emerges. Based on this causal assumption, participants should have had preference over MEET after viewing overlapping events. Instead participants parsed events, distinguished whether events overlapped or not, and weighted ends over beginnings. Despite the fact that there were always distractor events while the target events were unfolding, the results indicated that participants actively segmented the simultaneous events. For example, the

performance on START and FINISH was significantly above chance.

Participants tuned to the conceptual driven part of event temporal relations, i.e., the end states. Unlike the typical scripts used in the previous studies, perceivers could not predict the exact order of the events in the current study, however. An important question for future studies is the extent to which the previous knowledge of the simultaneous events has on event temporal relations. For example, would knowing that two events should start at the same time make people more likely to judge events as START even if they do not in fact actually start at exactly the same time? The parallel sessions at a conference are such events.

As in Experiment 1, we performed an analysis to see if the spatial locations of the end states in the animations may adversely affect the results. The proportions of the correct judgments on events that ended at one side of the animation frame were .28, whereas the proportions of the correct judgments on events ended at different sides of the frame were .29. When judging the event temporal relations, participants received sentences describing the target events and the locations where each target event ended. Knowing the goal location does not predict knowing the precise point of time an event ends. Nevertheless, the results of Experiment 2 cannot entirely rule out the possibility that describing events may give greater leverage to end states in constructing event temporal relations.

Many temporal relations bear similarity to each other. Given the importance of similarity in cognition (see Goldstone for a review, 1994), a question arises to how similarity between the event temporal relations might account for the results. Upon considering this question, there is one assumption that needs to be acknowledged: people segment events into meaningful units. Consider, for example, FINISH has greater surface structural similarities to both START and EQUAL than to other relations, because these three types of relations have one event contained

within the other event's temporal context. Compared with Experiment 1, the differences between the correct judgments made on EQUAL and FINISH decreased. Participants made more correct judgments on FINISH than on EQUAL, $t(39) = 1.66, p = .11$. The results were consistent with MDS solutions. In Experiment 1, START and FINISH had almost the same distance to EQUAL, whereas in Experiment 2, START was farther from EQUAL than FINISH. By the account of similarity alone, one could predict that people would tend to make temporal judgments based on surface structural similarity when the difficulty of event segmentation increases. The results did not provide strong support that similarity of events were the sole driving force.

Simple Recurrent Neural Network Judging Event Temporal Relations

Experiments 1 and 2 kept the amount of temporal overlap across different event temporal relations as close as possible. We could not entirely rule out the possibility that the similarities of event durations drive the confusion matrices. Therefore, we performed a simulation that systematically varied the amount of overlap between events to investigate this possibility. We chose to use a recurrent neural network in order to more closely parallel how human participants would experience the inputs of these two events as continuous time streams, where events begin and end at different points in time during the duration of an animation. If the boundaries of events and the subsequent distinction between overlap and non-overlap were not an important part of the representations, then the RNN performance would not be able to replicate human performance without being explicitly trained to do so.

Events

There were two inputs to the network, representing two generic events A and B. Input to the network was broken into 25 discrete time steps. Fig 4 presents an example of an OVERLAP relation used in the training of the network. Event A starts at time $t = 5$ and ends at time $t = 10$,

whereas event B starts at $t = 7$ and ends at $t = 15$. Each discrete time step represents the value fed into the RNN input unit for that time step. For example, at time step 5, x_A received an input of 1 while x_B receives an input of 0.

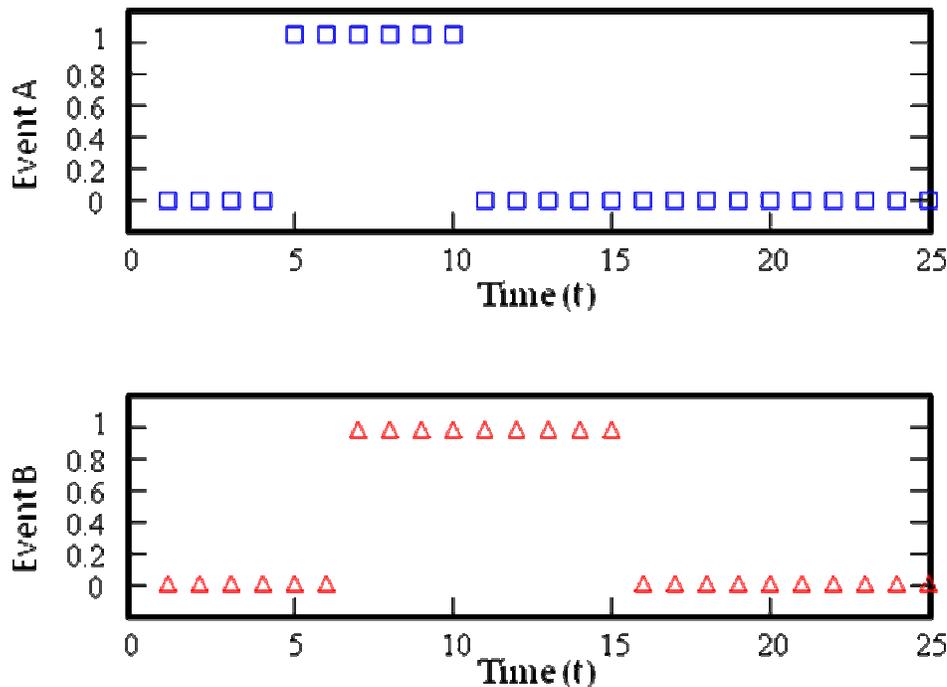


Fig 4.

Example of a crisp OVERLAP event input to RNN.

Each event had a crisp beginning and end, that is, the input had a value of 1 when the event was occurring and a value of 0 when the event was not occurring. The simulation varied the ratio of the event durations and the amount of overlap between events. Thus the presentation of the inputs to the recurrent neural network consisted of a 2×25 matrix of binary values that were fed as input to the network over a period of 25 discrete time steps. Event initiations and durations were generated randomly from a uniform probability distribution. That is, the OVERLAP relation shown in Fig. 4 was equally likely to have had an overlap of only a single time step, versus all points in time except the beginning and / or end point, versus any combination in between. The

same held for all other examples of trained and tested events. Since the events and temporal relations were generated with uniform probability, there was a wide range of examples of each event type, from the easy and obvious to the extremely difficult cases.

Network Structure

The network in the simulations consisted of two input units and seven output units. We used 50 units in the hidden layer. This number of hidden units was determined by empirical investigation as providing satisfactory results in the simulations (Fig. 5). We used a standard Elman simple recurrent neural network (RNN) structure. The Elman network is the simplest form of a RNN, with an input, hidden and output layer. Hidden layer activation at time t is computed based on the activation of the input units and a copy of the hidden units activation in a previous time step. Output is computed for time t based on the activation of the hidden layer units at time t . Recurrence occurs because a copy of the activation of the hidden unit layer in the previous time step is re-fed in as input.

The networks were implemented using the Stuttgart Neural Network Simulator (SNNS) version 4.2 (1991). All weights were subject to learning by a variation of backpropagation training, known as backpropagation through time, using the SNNS JE_BP_Momentum learning function (Jordan/Elman network backpropagation through time learning function with momentum). All feedforward weights were initialized to random values between -1.0 and 1.0 before training. The weights from the hidden units to the context units were fixed with value 1.0 and were not modified by learning. A learning rate of 0.1 and momentum term of 0.5 were used for all training results reported. Backpropagation through time is a variation of standard feed forward backpropagation learning that unrolls the feedback connections of recurrent neural network so that standard backpropagation learning can still be used to train such networks.

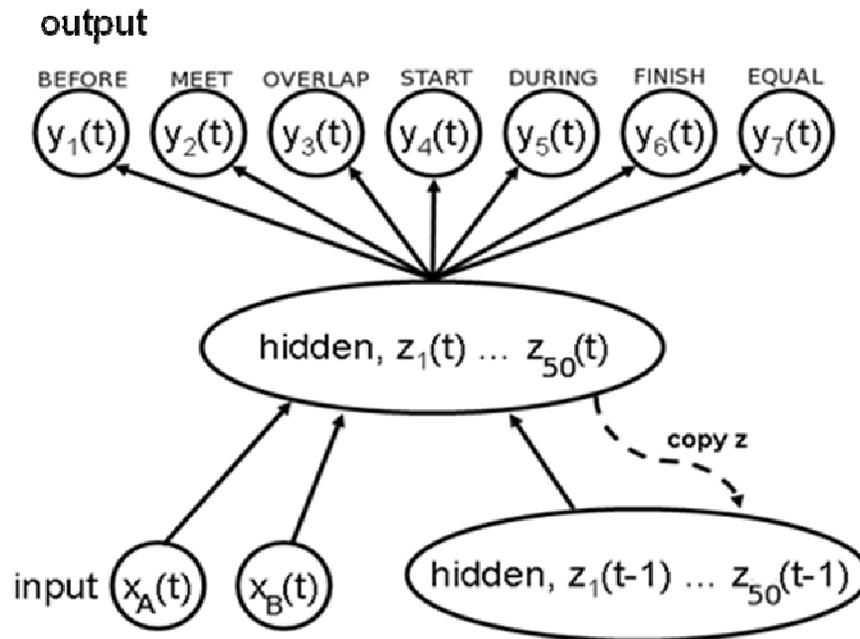


Fig 5.

Structure of the Elman simple recurrent neural network used in the experiments. Inputs consisted of two event inputs (x_A and x_B) and seven outputs (y_1 through y_7), representing one each of Allen's seven temporal relations. Recurrence occurs by copying the activation of the hidden units from a previous time step and (re)feeding this activation back into the network. All solid arrows represent plastic connections whose weights are modified using standard backpropagation through time during training.

Training

The training set consisted of 1000 randomly generated event temporal relations, similar to the one shown in Fig 4, but uniformly distributed over Allen's seven temporal relations. We used seven output units in the recurrent neural networks, each unit trained to respond to the occurrence of one of Allen's temporal relations of the two input event streams. The networks were trained to generate an output of 1 for the indicated relationship as soon as both events were completed. For example, if two events had an OVERLAP relation as shown in Fig 4, the node in the network being trained to represent OVERLAP relations (y_3) was taught to generate 1 at time step 16, and keep generating 1 until time step 25. Fig 6 shows an example of the responses of a network after being

trained. In the left hand side of Fig 6 an example of an incorrect judgment is shown. The two events presented to the network in the left figure represented a BEFORE relation. In the figure, the horizontal axis represents the time steps of events in the simulation, from 1 to 25, and the vertical axis plots the activations of each of the seven output units in the network. As can be seen in the left figure, the MEET and BEFORE nodes both respond to the input events. However, the MEET node responds much more strongly, even though the events were actually a BEFORE relation.

In the simulations presented in this paper, we determined the winning judgment by taking the average activation of the output unit over the last 5 time steps (from time step 21 to 25). The average activation of the MEET output unit was about 0.9 over the last 5 time steps, while only about 0.15 for the BEFORE output unit. The right hand graph of Figure 5 shows an example of a correct judgment. Here a START relation was presented to the network. In this example four nodes responded somewhat to the START events. However, the START node responded with the highest average activation over the last 5 time steps, so it would be declared as the winning judgment.

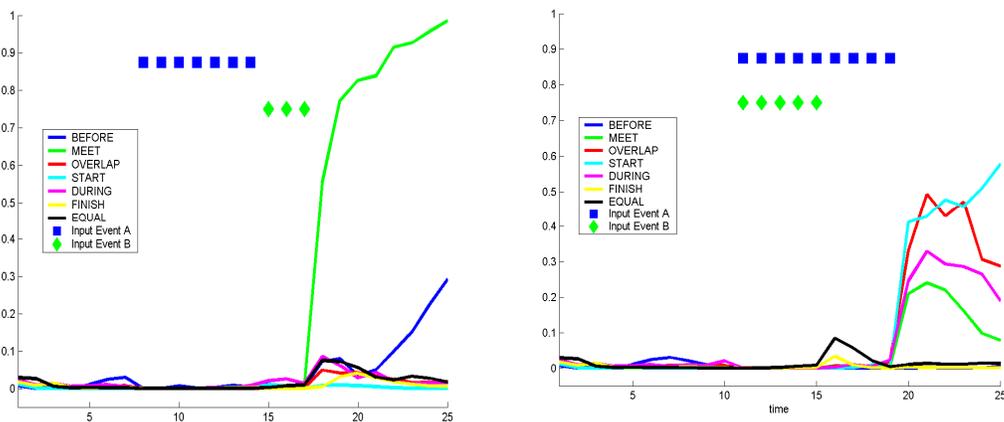


Fig 6.

Examples of an incorrect judgment (left) and a correct judgment (right).

Results

After training, weights were fixed, and the trained network was tested on a new set of 10,000 randomly generated generic events, again spread evenly among the seven event temporal relations. Thus we used holdout cross-validation with 1000 randomly created training examples, and 10,000 randomly generated testing examples. We determined the performance of the network using a winner-take-all strategy. The relationship represented by the winning node was selected as the judgment of the event temporal relation experienced by the network. Table 6 shows the network performance on the test event set. The network performance was then compared with the human performance results on Experiment 1 shown in Table 4 and showed a significant correlation, $r(49) = 0.98, p < .001$. The pattern of errors produced by the network also showed a very high similarity to those produced by the human participants, $r(42) = .50, p < .05$.

Table 6. *Proportions of network judgments on events*

| Events in the World | Network Judgment | | | | | | |
|---------------------|------------------|------|---------|-------|--------|--------|-------|
| | BEFORE | MEET | OVERLAP | START | DURING | FINISH | EQUAL |
| BEFORE | .882 | .042 | .024 | .021 | .011 | .009 | .011 |
| MEET | .046 | .830 | .065 | .023 | .014 | .010 | .012 |
| OVERLAP | 0 | .011 | .779 | .166 | .040 | .002 | .001 |
| START | .003 | .006 | .118 | .760 | .104 | .005 | .005 |
| DURING | .003 | .007 | .078 | .279 | .625 | .003 | .006 |
| FINISH | .004 | .004 | .022 | .009 | .002 | .678 | .280 |
| EQUAL | .003 | .006 | .008 | .008 | .005 | .056 | .913 |

When comparing the network performance to the hypotheses, the network error patterns were very similar to the end state bias predictions after taking into account the overlap heuristic, r

(42) = .73, $p < .001$, and were also similar to the overlap heuristic alone, $r(42) = .35, p < .05$. The network error patterns were not correlated with the beginning state predictions. The MDS solution maintained the basic patterns from the previous experiments.

Discussion

The RNN network is capable of forming correct discriminations of the temporal relations presented to it. The fact that a RNN can succeed at this task was not obvious a priori because the judgment requires remembering the relative timing of beginning and end states, as well as temporal overlaps, between two input streams that could extend arbitrarily far back in time as experienced by the network. The overall proportion of errors made by the networks were .219 (SD = .10), similar to the results of human participants in Experiment 1a where the overall error rate was .212 (SD = .12). The network demonstrated the formation of an overlap distinction, despite the fact that the network was not explicitly trained to do so. The RNN simulation performed very well on BEFORE and MEET, and when it was incorrect, it tended to confuse these types of event relations with one another, although MEET was also understandably confused with OVERLAP at times. Also, OVERLAP, START, and DURING relations appeared to be confused among one another by the network, as with humans, though FINISH and EQUAL were much more similar within this overlapping group among human participants in Experiment 1 than the network simulation. This tendency to confuse the non-overlapping event relations with each other, and likewise the overlapping event relations with one another, is why we infer the formation of a general high-level distinction of overlapping versus non-overlapping events.

The experiences of each type of event boundary seemed to have left imprints on the neural network's representations and pushed the organization of the network's dynamics into distinct attractors. That is to say, there is evidence that the network used event boundaries where events

overlapped or broke apart from each other in judging event temporal relations. The network did not simply key in on one simple distinction, such as remembering the relative end state of the events (else it would build categories such as those events that end at the same time like FINISH and EQUAL, and those events that do not). The network used multiple cues in combination to form several higher-level attractors that it used to distinguish the various event temporal relations in a similar fashion to human participants.

Many factors, such as the passage of time and closeness of event boundaries, can easily collapse unique trajectories into similar “remembrances” of the temporal ordering of the experienced events, and thus were reflected as patterns of confusion among the event categories in ways similar to human confusions. As a result, the non-overlapping relations BEFORE and MEET form a self-similar group, while OVERLAP, START, and DURING form another, with possibly FINISH and EQUAL forming a third broad category in the network simulation. The further categorizations of the overlapping relations appear, in the neural network case, to be mainly broken down depending on the feature of whether the two events ended simultaneously (e.g. as in a FINISH or EQUAL relation), or whether they ended at different times (OVERLAP, START, DURING). The overlap distinction was emergent from the spontaneous event segmentation task the network was trained to perform. Thus, the specific shape of the attractor was not controlled directly, but it self-organized according to the nature of the event temporal relations it experienced while being trained.

All decisions made by the network must be based on local cues. From the networks point of view, at each time step there were only three possibilities: (a) nothing was happening, (b) event A (or B) was occurring but not the other, or (c) both events were occurring. The learning process involves shaping the attractor space such that from these purely local features emerge attractors

that capture the history of the event perception, and thus allow a recall of which relation occurred depending on where the system ends up in the phase space. The network exploited the sensory characteristics in the input events and used them to guide perceptual segmentation. This exploitation of event breakpoints in the environment, a process of self-organization, reminds us that intelligent agents exploit their niche for their purposes, mostly without knowing that they do so.

Though the network did not exclusively use the end state of the events in making its judgments, it did demonstrate a significant bias toward the end state and appeared to be more influenced by recent event properties held in memory. This could explain some of the differences in the network performance versus human performance such as FINISH and EQUAL confusions. The relatively short historical window is a known limitation of simple RNN type networks. Basically, the dynamics of the recurrent connections of such a network theoretically allows events far in the past to effect the current dynamics of the network. However, in practice, the dynamic influence into the future is a rapidly decreasing nonlinear effect, and in practical terms such networks have problems remembering and judging properties that occur more than 5 or 6 discrete time steps in the past.

General Discussion

The hypothesized differential status of event boundaries, which guide perceiving, remembering, and planning events, shapes the construction of event temporal relations. It is not the case where people only use one type of event boundary (e.g., end states) to construct event temporal relations. The stimulus events in the current study were clear cut. Therefore, the psychological and physical event boundaries tend to have a greater degree of alignment. However, for events that have fuzzy boundaries, the locations where people place event breakpoints will be

even more important for constructing temporal relations.

In Experiment 1, we presented simple animations that have crisp and clear cut event boundaries. Participants selected one of Allen's seven event temporal relations after watching each animation. Instead of primarily using either beginnings or ends as an anchor in constructing event temporal relations, the perceptual segmentations of two tracks of events appeared to have impacted event temporal relations. For example, the event segmentation would suggest that people perceive OVERLAP animations to have three visual events: (a) the event of one fish beginning to swim alone, (b) the second event of two fish swimming simultaneously, and (c) the third event of the late start fish swimming alone. In the pilot study we performed where participants spontaneously described events, the number of clauses for the following types of animations matched with what event segmentations would suggest: BEFORE ($M = 3.52$), MEET ($M = 3.36$), OVERLAP ($M = 3.34$), and DURING ($M = 3.17$). Similarly, the number of clauses for the following animations matched with what event segmentation would suggest: START ($M = 2.62$), FINISH ($M = 2.72$). In Experiment 2, we presented animations that are embedded in schemata and that have target events accompanied by simultaneous distractor events, which could force participants to resort to the similarities of event overlaps. The overlap heuristic hypothesis had the most predictive power over judging event temporal relations.

The RNN simulation showed that there is a natural tendency to self-organize event temporal relations into two broad categories, overlapping versus non-overlapping type of events. The RNN simulation appears to have significant bias towards remembering end states, thus it performed much more poorly on START relations, and tended to confuse FINISH with EQUAL relations much more than human participants. This would appear to be a by-product of the network architecture and not any artifact of the structure or set up of our input to the simulation. If

human judgments were more similar to the RNN model, it would indicate more of an emphasis on remembering end states in humans and imply that humans have similar short term memory effects when remembering and judging temporal relations. However, the differences between model and human performance seem to imply that humans are doing more than simply remembering the immediate end state of the events. Future research needs to investigate what might give advantage to the end states: Is it because end states occur more closely to the retrieval of event temporal relations as compared with the beginning states? Or is it because end states across different instantiations of events are more similar?

An interesting aspect of self-organization not directly addressed in this study is the effect that systematic variation on a continuum might have on event temporal relation judgments. When training the network simulation, the network received examples of all event temporal relation types in no particular (random) ordering. It is known that, in the domain of speech perception (Tuller et. al. 1994, Tuller 2003, Kelso 1995), systematic variation of the event interval between speech sounds, in a stepwise fashion, can induce hysteresis or assimilation effects. People are much more likely to continue identifying a word with a gap in pronunciation (such as between the *s* in pronouncing say versus stay) as their original perception as the gap is systematically increased or decreased. Humans may experience similar hysteresis effects when learning about and judging event temporal relations. For example, referees or spectators of fast paced sporting events may be biased in their perception of the relative temporal ordering of significant events after experiencing similar situations during the course of the event. In our human experiments, the relative timing of events was held relatively constant, thus minimizing any possible such hysteresis effects during temporal judgments. Some differences in the simulation performance could possibly be the result of these types of effects, since the network received much more variations in temporal timings

among the trained and tested events it experienced.

Significant changes take place at event boundaries, whether they be attention, perception, or conception updates. The current research showed how event boundaries affect the subsequent formation of event temporal relations. Our research suggested that the temporal representations of events are richer than previously assumed (Block, 1990). That is, people segment simultaneous events and use these event boundaries in constructing event temporal relations rather than merely constructing sketchy event temporal relations. Further studies are needed to investigate the conditions under which people could perform simultaneous segmentations and how the task environments affect the locations of the event breakpoints.

References

- Abbott, V., Black, J. H., & Smith, E. E. (1985). The representation of scripts in memory. *Journal of Memory and Language*, *24*, 179-199.
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1-12.
- Allen, J. F. (1984). Towards a general theory of action and time. *Artificial Intelligence*, *23*, 123-154.
- Allen, J. F. (1991). Time and time again: The many ways to represent time. *International Journal of Intelligent Systems*, *6*, 341-355.
- Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1-to-2-year-olds' recall of events. *Developmental Psychology*, *25*, 197-206.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.
- Block, R. A. (1990). Models of psychological time. In R. A. Block (Eds.), *Cognitive models of psychological time* (pp. 1-35). Hillsdale, NJ: Erlbaum.
- Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *16*, 190-195.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*, 395-429.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177-220.
- Chessman, P. (1985). In defense of probability. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence* (pp.1002-1009). Los Angeles, CA: Morgan Kaufmann.

- Chessman, P. (1988). An inquiry into computer understanding. *Computational Intelligence*, 4, 58-66.
- Davidson, D. (1996). Events as particulars. In R. Casati & A. C. Vazi (Eds.), *Events* (pp.99-107). Aldershot, England: Dartmouth (Reprinted from *Nous*, 4, pp. 25-32, 1970).
- Davis, L. H. (1970). Individuation of actions. *Journal of Philosophy*, 68, 520-530.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, 113, 44-66.
- Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.
- Gennari, S. P. (2004). Temporal references and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 877-890.
- Golding, J. M., Magliano, J., & Hemphill, D. (1992). When: A model for answering “when” questions about future events. In T. W. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 213-282). Hillsdale, NJ: Earlbaum.
- Goldman, A. (1970). *A Theory of Human Action*. Princeton, NJ: Princeton University Press.
- Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52, 125-157.
- Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of statements about story worlds that deviate from normal conceptions of time: What is true about *Einstein's Dreams*? *Cognitive Psychology*, 35, 246-301.
- Hacker, P. M. S. (1996). Events and objects in space and time. In R. Casati & A. C. Vazi (Eds.), *Events* (pp.99-107). Aldershot, England: Dartmouth (Reprinted from *Mind*, I, pp. 1-19, 1982).
- Jarvis, B. G. (2000). MediaLab research software (Version 2000) [Computer software]. New York: Empirisoft.
- Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (Tech, Rep. No. 8604). University of California, San Diego, Institute for Cognitive Science.

- Kelso, J.A.S. (1995). *Dynamic Patterns: The self-organization of brain and behavior*.
Cambridge, MA: The MIT Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: Chicago University Press.
- Lakusta, L. (2006). Source and goal asymmetry in non-linguistic motion event representations.
Dissertation Abstract International, 66(12), 6944. (UMI No. 3197182).
- Lichtenstein, E. D., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology*, 12, 412-445.
- Linton, M. (1975). Memory for real-world events. In D. A. Norman, & Rumelhart, D. E. (Eds.),
Explorations in cognition (pp. 376-404). San Francisco, CA: Freeman.
- Loftus, E. F., Schooler, J. W., Boone, S. M., & Kline, D. (1987). Time went by so slowly:
Overestimation of event duration by males and females. *Applied Cognitive Psychology*, 1,
3-13.
- Michotte, A. E. (1963). *The perception of causality* (T. R. Miles & E. Mile, Trans). New York:
Basic Books. (Original work published 1946)
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity
for processing information. *Psychological Review*, 63, 81-97.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New
York: Holt, Rinehart & Winston.
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for
social and physical events. *Journal of Personality and Social Psychology*, 67, 949-971.
- Mourelatos, A. P. D. (1978). Events, processes, and states. *Linguistics and Philosophy*, 2,
415-434.
- Nelson, K. (Eds.). (1986). *Event knowledge: Structure and function in development*. Hillsdale, NJ:
Erlbaum.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of
Personality and Social Psychology*, 28, 28-38.

- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
- Regier, T., & Zheng, M. (2003). An attentional constraint on spatial meaning. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Reynolds, J., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*, 613-643.
- Rumelhart, D.E., & Norman, D. A. (1985). Representation of knowledge. In A. M. Aitkenhead & J. M. Slack (Eds.), *Issues in cognitive modeling* (pp. 15-62). London: Erlbaum Associates.
- Schank, R. C. (1999). *Dynamic memory revisited*. New York: Cambridge University Press.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- Scholl, B. J., & Nakayama, K. (2002). Causal capture: Contextual effects on the perception of collision events. *Psychological Science*, *13*, 493-498.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379-423. Continued in following volume.
- Thomson, J. J. (1971). The time of killing. *Journal of Philosophy*, *68*, 115-132.
- Tuller, B. (2003). *Computational models in speech perception*. *Journal of Phonetics*, *31*, 503-507.
- Tuller, B., Case, P., Ding, M. & Kelso, J.A.S. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1-14.
- Wolff, P. (2003). Direct causation in the linguistic coding and individuation of causal events. *Cognition*, *88*, 1-48.
- Yamada, Y., & Kawahara, J. (2007). Dividing attention between two different categories and locations in rapid serial visual presentations. *Perception and Psychophysics*, *69*, 1219-1230.

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

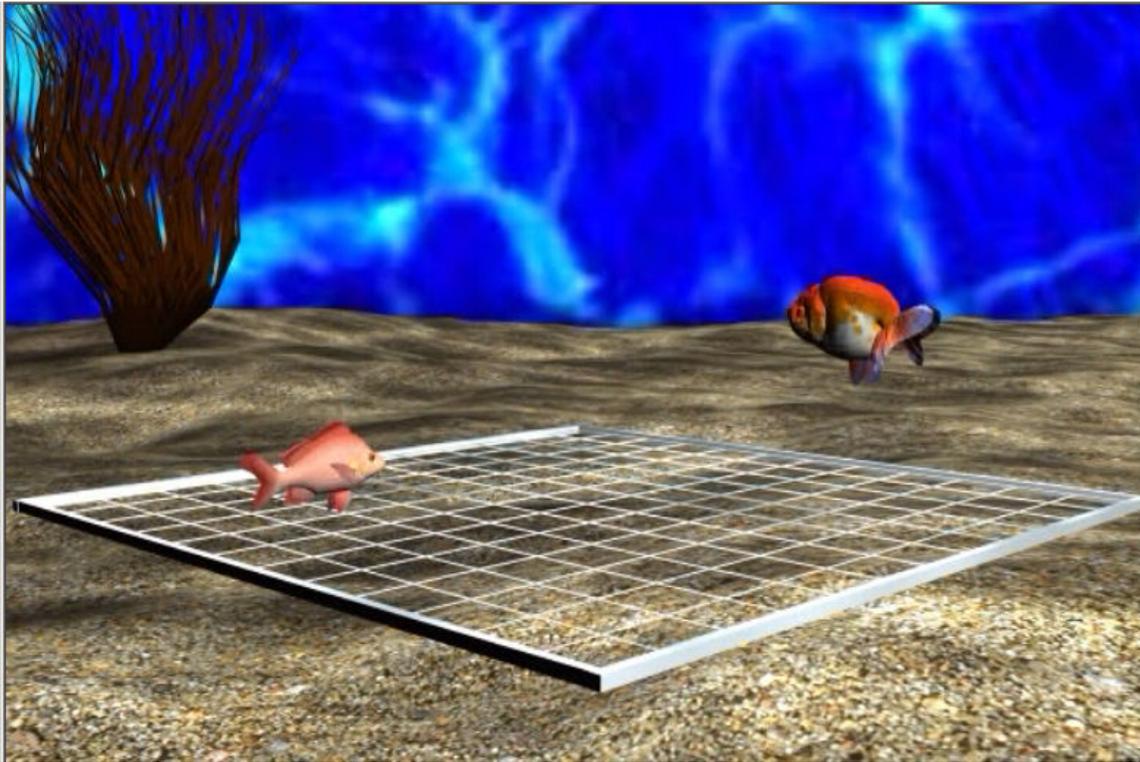
Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3-21.

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

Author Note

This research originated from Shulan Lu's dissertation work at University of Memphis. We are grateful to Phillip Wolff, Danielle McNamara, and Donald Franceschetti serving as committee members, providing feedback, and stimulating discussions on event representations. Part of this research was supported by a Research Enhancement Grant from Texas A & M University – Commerce. Part of this research was also supported by a Texas Advanced Research Program (ARP) grant from the Texas Higher Education Coordinating Board. We thank Srinivas Achunala for constructing the 3d animations, Richard Brooks Hansen, Lonnie Wakefield, Stephanie Coe, and Dorothy Presbury for assistance at the various stage of this work. We thank Jeff Zacks for discussions and feedback on the manuscripts. Correspondence concerning this article should be addressed to Shulan Lu, Department of Psychology, Texas A&M University-Commerce, TX, 75429-3011, Shulan_Lu@tamu-commerce.edu.

Appendix 1: A snapshot of a sample animation used in Experiment 1



Appendix 2: A snapshot of a sample animation used in Experiment 2

