# Perceiving, Remembering and Communicating Structure in Events

Jeff Zacks, Barbara Tversky, and Gowri Iyer

Stanford University

October, 1999

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

## Abstract

How do people perceive routine events such as making a bed as they unfold in time? Research on knowledge structures suggests that people conceive of events as goal-directed partonomic hierarchies. Here, participants segmented videos of events into coarse and fine units on separate viewings; some described the activity of each unit as well. Both segmentation and descriptions support the hierarchical bias hypothesis in event perception: observers spontaneously encoded the events in terms of partonomic hierarchies. Hierarchical organization was strengthened by simultaneous description, and to a weaker extent, by familiarity. Describing from memory rather than perception yielded fewer units but did not alter the qualitative nature of the descriptions. Although the descriptions were telegraphic and without communicative intent, their hierarchical structure was evident to naive readers. The data suggest that cognitive schemata mediate between perceptual and functional information about events, and indicate that these knowledge structures may be organized around object/action units.

## From Planning, Reading, and Remembering to Perceiving

Events unfold in time, from the mundane making of a bed to the momentous making of a war. Observers can in principle use temporal structure to respond appropriately in real-time activity, to plan future action, to remember the past, and to coordinate with others. In particular, events can be decomposed into temporal parts, just as objects can be decomposed into spatial parts, and these parts can be related to each other. Are people sensitive to this structure, and if so what governs the relationships they perceive?

The ability to identify the parts of events and their relationships constitutes a distinct perceptual process, which we will call event structure perception. An event is defined to be a segment of time at a given location that is perceived by an observer to have a beginning and an end. In particular, we are concerned here with the perception of events in mundane, goal-directed activities. These are activities that our experimental participants might encounter on any given day, that generally have durations of several minutes, and that are performed by people with particular goals in mind. This paper describes a series of experiments that systematically explore the perceptual structure of events (Experiment 1), the relationship of familiarity and expertise to that structure (Experiments 2 and 3), the role of event structure in memory for events (Experiment 4) and in communication (Experiment 5). Together, the results strongly suggest that observers are biased to perceive ongoing activity in terms of discrete events organized hierarchically by "part-of" relationships. This disposition is revealed in encoding of ongoing events, memory for past events, and discourse about events.

These experiments were conducted to explicate the relationships between the on-line perception of events and off-line conceptions of events. The latter are important for planning, understanding narratives, and remembering past events—and have been examined extensively through studies of those processes (Zacks & Tversky, submitted). It is to those conceptions that we now turn.

Action Planning

People often describe plans in terms of discrete steps that are related by an overarching structure. To explain to someone how to get from downtown Palo Alto, CA to the Golden Gate Bridge, one might begin: "Get on highway 101 going north. Take it to San Francisco. The highway will end in city streets. Follow the 101 signs." If the directee were unfamiliar with Palo Alto, the first step would have to be expanded: "Find University Avenue, the main drive. Drive East on University Avenue away from the University. As University Avenue leaves Palo Alto, take the entrance for highway 101 North." That is, to further explain a given step, one breaks it down into a series of sub-steps.

Newell and Simon (1972) argued that planning proceeds in the same fashion. In their General Problem Solver (GPS), a plan begins as a high-level goal (get to the Golden Gate Bridge). Subgoals are identified whose satisfaction will lead to the satisfaction of the original goal (get to the highway, take it to San Francisco, etc.). This process proceeds recursively until each of the subgoals in the plan can be achieved by a behavioral primitive. The resulting plan has the form of a hierarchy. This hierarchical structure explains how we talk about plans when explaining them to others: "Episodes, since they are tied to goals, can be hierarchical, with one episode embedded in another" (Newell & Simon, 1972, p. 480). "Unrolling" a GPS plan into actions transpiring over time gives rise to a

hierarchical structure. Actions designed to satisfy goals at a given level are subdivided in time into sub-actions designed to satisfy subgoals. The goal-subgoal relationship is cached out as a part-subpart relationship, leading to a partonomic hierarchy. Vallacher and Wegner (1987) argue that goals at multiple levels can condition action even as it is being performed, and that action tends to be controlled from the highest level available in the goal hierarchy.

Narrative Comprehension

Narratives are discourses that describe a set of actions. If actions can be thought about in terms of hierarchical part structures, it stands to reason that people apply these structures to understanding narratives. Research on schemata for stories, story grammars, and scripts abounds with support for this intuition (e.g., Bower, 1982; Mandler & Johnson, 1977; Schank & Abelson, 1977; Thorndyke, 1977; Trabasso & Stein, 1994). Rumelhart (1977) argued that we understand stories by recourse to internalized schemata that have a partonomic structure organized around goals and subgoals. In his model, comprehension of a story corresponds to matching the text to a schema. Summarization can be modeled as pruning of the hierarchical representation to higher levels. Finally, recall is modeled by a two-step process. First, stored traces from reading the story are located and activate the appropriate schema. Then, the schema and traces together are used to reconstruct the details of what happened in the story. Valid story schemata correspond to a grammar which defines legal partonomic relationships (Rumelhart, 1975).

Schank, Abelson, and their colleagues formulated a set of computational models based on the related notion of a script (Schank & Abelson, 1977). Scripts are a particular implementation of event schemata designed to account for

understanding of goal-directed activities that recur in everyday life. Like Rumelhart's (1977) schemata, scripts are organized as partonomic hierarchies in which a script consists of a set of scenes, and each scene contains a set of actions. Computer simulations based on scripts have achieved a respectable level of success in understanding newspaper stories and simple narratives (Schank & Abelson, 1977). Furthermore, a number of phenomena in the comprehension and memory of narratives support the psychological reality of the script concept (Abelson, 1981). In particular, the literature on text comprehension and memory strongly supports the view that hierarchical representations of events play a part in understanding narratives. During reading, larger pauses occur across higher-level event boundaries (Abbott, Black, & Smith, 1985). Short narratives are read more quickly if the higher-level structure is established early in the text (Foss & Bower, 1986).

Memory for Events

Hierarchical organization seems to influence not just on-line processing of narrative text, but also memory for them. Participants are slower to answer questions about a text when the questions require integrating information across higher-level event boundaries; this is true even when distance in the text is controlled (Foss & Bower, 1986; but see Franklin & Bower, 1988). People sometimes falsely recognized action statements that were omitted from a story but were implied by its script (Bower, Black, & Turner, 1979), and these inferences tended to generalize upward in the hierarchy (Abbott et al., 1985). Investigators have explicitly tied these results to the theories of planning described previously, by noting that hierarchies are tied to plans (Bower, 1982) and are useful for organizing planning (Abbott et al., 1985).

Similar results obtain for memory of videotapes, a stimulus type that is closer to "live" events than narrative texts. Recall of videotapes of human activity is characterized by a hierarchical pattern of recall. Memory for actions that are relevant to the event schema is better than memory for schema-irrelevant actions. As with memory for texts, the order of sub-events tends to revert with time to the schema-normal order (Lichtenstein & Brewer, 1980). As with texts, activation of a schema can lead to false recognition of actions implied by that schema. Further, the same action is better recognized and better recalled when it is part of an activated event schema than when it is not, and recall for details within an event segment tends to be all-or-none (Brewer & Dupree, 1983).

Hierarchical patterns in recall are also seen in autobiographical memory, that is, in memory for the narrative of one's own life. Once a given episode has been activated in memory, its' sub-parts are more available (Anderson & Conway, 1993). Over time, presumably with deterioration of specific information, memory for autobiographical events shows an increasing influence of schemata on recall (Barsalou, 1988).

Developmentally, event structure influences recall from an early age. Hierarchical patterns of recall and effects of goals on memory for activity have been found for stories in 4-6-year-olds (e.g, Hudson, 1988; Nelson & Gruendel, 1986; van den Broek, Lorch, & Thurlow, 1996), and for simple events in infants as young as 15 months (e.g., Bauer & Mandler, 1989; Travis, 1997).

Implications for Perception

A general picture emerges in which activity is thought of in terms of hierarchically organized relations among "chunks" at different temporal grains. This structure influences how people think about planning activity, how they

comprehend and remember texts that describe activity, how they remember those texts (or videotapes of similar narratives), and how they remember the activity in their own lives. Does it influence perception as well? One reasonable hypothesis is that the same representations that support conceptions of events will also play a role in event perception. This leads to a prediction: People will be spontaneously disposed to actively encode ongoing activity in terms of a hierarchical part structure. We will call this the hierarchical bias hypothesis.

There are also several compelling alternatives. One possibility is that temporal relationships of this sort may not be directly given in perception. That is, observers may be able to extract arbitrary temporal structure from activity on request, but they do not spontaneously track it, let alone track relationships across temporal grains. Another possibility is observers do spontaneously track temporal structure, but the structures to which they are sensitive are not hierarchical. For example, coarse-level segmentation might be performed on the basis of goals and fine-level segmentation on the basis of perceived changes in physical activity. The conceptual bases for segmentation might not correspond, yielding unaligned coarse and fine units.

We know of no empirical research aimed at examining this question directly. However, one line of research on the relationship between event segmentation and social-personality attribution applies—and to the extent it does, it argues against the hierarchical bias hypothesis.

Newtson (1973) developed a technique in which participants segment ongoing activity while watching it on videotape by pressing a key to mark "natural and meaningful" unit boundaries. In one experiment, the grain at which participants segmented the activity was manipulated between subjects:

One group was asked to make the largest natural and meaningful units; the other group the smallest. Newtson found that points in the activity that participants in the large-unit condition tended to agree were unit boundaries also tended to be boundaries for the small-unit group. Conversely, points in the activity that were not marked as unit boundaries by the small-unit participants also tended not to be marked as unit boundaries by the large-unit participants. Based on this result, Newtson concluded that participants in the small-unit condition identified units that were subdivisions of those identified by the large-unit participants.

One might take these data as arguing for the hierarchical bias effect. However, Ebbesen (1980) has argued against this view based on a characterization of the segmentation task as a "secondary" task that does not reflect natural encoding processes. In one experiment, Cohen and Ebbesen (1979) asked participants to segment a videotape using Newtson's (1973) method, under instructions to either learn the task being performed by the actor or form an impression of the actor's personality. They found that under impression-formation instructions participants produced larger units than under task-learning instructions. However, they reported poor within-participant agreement on the location of unit boundaries in the two conditions. Based on this result, Ebbesen concluded that unit boundaries "do not appear to be hierarchically structured, as Newtson (1973) suggested" (Ebbesen, 1980, p. 188). Drawing conclusions from these studies is difficult, given their differing results. In addition, in Cohen and Ebbesen's (1979) study, no formal test of unit-location agreement or lack thereof was reported. These apparently conflicting patterns have no serious implications for the primary aims of the studies that generated

them (which were concerned with how participants vary their encoding patterns and how these relate to patterns of attribution). However, this uncertainty regarding the possible hierarchical structure of event boundaries during encoding makes a case for careful study of the perceptual encoding process.

Moreover, even if one accepts the hierarchical segmentation pattern, one can reject the conclusion that this reflects a cognitive representation of hierarchically organized events. In fact, this is exactly the position Newtson has taken in later work, in which he argues that patterns of event segmentation are best interpreted in terms of dynamical systems, as reflecting the topology of a system which includes the observer as well as the activity being observed (Newtson, 1993; Newtson, Hairfield, Bloomingdale, & Cutino, 1987). If hierarchically organized cognitive representations are playing a role in perceptual encoding, one should be able to observe more than simply patterns in encoding behavior. Segmentation patterns should make rich contact with downstream processes such as language and memory, and should be influenced by prior experience.

The experiments presented here were designed to test the hierarchical bias hypothesis, to examine the structure of event segmentation across time scales, and to relate perceptual processing of temporal information to other aspects of cognition. The first goal of these experiments was to provide a stringent within-subjects test of the hypothesis that observers would segment events in terms of a partonomic hierarchy. The second was to examine the influence of prior experience with a particular activity on event segmentation. The third goal was to characterize the relationships of higher-level cognitive operations such as language and memory to event structure perception. The

final objective was to examine how people can use hierarchical organization to communicate with others about activity.

## Experiment 1: Perception of Event Structure

To the extent that the mind makes use of hierarchically organized schemata for events, and these schemata influence perception, one should observe a bias to encode activity in terms of partonomic hierarchies. However, the small amount of relevant research is in conflict (Cohen & Ebbesen, 1979; Newtson, 1973; Newtson & Engquist, 1976). The first major goal of this experiment was to provide a direct test of the hypothesis that observers spontaneously segment activity such that it corresponds to a partonomic hierarchy, i.e. to test the hierarchical bias hypothesis.

Second, we wanted to test the hypothesis that descriptions of ongoing activity would reflect the same structure, and elucidate its origins. Hierarchical segmentation is not sufficient to establish on what basis observers organize activity. Language analysis may be particularly valuable in this regard, particularly given work in linguistics arguing that language structure reflects an underlying cognitive structure for events (Goldberg, 1995; Levin, 1993; Moens & Steedman, 1988; Narayanan, 1997; Pustejovsky, 1991; Talmy, 1975).

Finally, we wanted to test a pair of hypotheses about the factors that mediate the influence of structured representations on event perception. To the extent that language and event representations are tightly integrated online, linguistic representations of events should activate, as well as be activated by, perceptual representations. Producing an adequate description of ongoing action may require making connections across temporal grains—even if the

description is restricted to one temporal grain, as it was here. Talking about activity may require activation of representation of information about the goals and plans of the actor(s). This leads to the prediction that talking about activity as it happens should increase the tendency to organize it in terms of relevant event schemata.

Event schemata can only be present for activities with which one has had some sort of prior experience, which leads to the prediction that observers should show a greater tendency to segment activity hierarchically for familiar activities than for unfamiliar activities.

To test these hypotheses and explore their consequences, we adapted Newtson's (1973) segmentation procedure and applied it to the perception of four everyday activities. Participants viewed videotapes of the activities and were asked to segment and describe them while watching. A control group only performed the segmentation. Each participant segmented each activity twice, in counterbalanced order, once providing coarse units and once providing fine units. Segment boundaries from these two viewings were compared to provide an estimate of the degree to which the viewer was spontaneously encoding the activity hierarchically. The prediction of the hierarchical bias hypothesis is that, for a given participant, each coarse-unit boundary for a given activity would tend to fall closer to some fine-unit boundary for that activity than predicted by chance. Furthermore, it was predicted that variations in the syntactic and semantic features of the language used to describe the activity would correlate with each participant's pattern of segmentation.

## Method

### Participants

A total of 40 Stanford University undergraduates participated in this experiment to partially fulfill a course requirement. Three additional participants were run, but their data were unusable due to technical difficulties, so they were replaced.

### Selecting Activities for Study

In preparation for selecting activities for the current research, ratings of frequency, familiarity, and knowledge of steps were obtained for 45 everyday activities. These norms are described in Appendix A, and included ratings of frequency of performance, familiarity, and knowledge of steps. From the 45 activities, two were selected that were rated low on all three scales ("assembling a saxophone" and "fertilizing houseplants"), and 2 that were rated high on all three scales ("washing dishes" and "making a bed"). Because the three scales were highly correlated (see Appendix A), we will refer to these activities as "unfamiliar" and "familiar" (respectively). As Figure 1 shows, both unfamiliar activities were much lower on all three ratings than both familiar activities. These four activities were used in all the experiments described in this report.

---

Insert Figure 1 about here

---

### Stimulus Films

For each of the four activities selected from the norms, we constructed a script consisting of twelve discrete steps (see Appendix B). The scripts were simply lists of twelve steps for the actors to perform, written in order to

encourage similar performances by the two actors. (By constraining the performances of the actors in this fashion, we hoped to be able to make quantitative comparisons across videotapes of the same activity. This proved infeasible due to substantial timing differences between actors.) No relationships were established between the steps in the list other than their serial order, nor were any such relationships discussed with the actors during filming. These precautions were taken to avoid building the presence of hierarchical structure into the stimuli. Two actors (one male, one female) performed each of the activities in accordance with the script. Each performance took place in a different location. Performances were recorded with a Hi-8 videotape camera and dubbed to VHS. The video camera was placed in a fixed head-height position, attempting to simulate the viewpoint of an observer in the room. Each activity was recorded as a single take, with no cuts, pans or zooms, to minimize the effects of cinematic conventions on participants' perceptions. The resulting tapes ranged in length from 244 to 640 seconds. Also, a sample tape was made with a third actor (female) and another activity (ironing a shirt).

Procedure

Participants were run individually. On entering the laboratory, each participant in this study was seated in front of a television, near a computer keyboard and a tape recorder. Participants were told that they would be shown a series of short videotapes and were instructed to tap the space bar on the keyboard "when, in your judgment, one unit ends and another begins." The 32 participants in the Describe group were then told: "Each time, after you press the space bar, say for the tape recorder what happened." For the 8 participants in the Silent group, the instruction to describe the activity after each tap was

omitted. The instructions made clear that they should tap exactly when they believed one unit ends and another begins, not in the middle.

Half of the participants in each group were instructed to "mark off the behavior of the person you'll be seeing into the smallest units that seem natural and meaningful to you." The other half was instructed to "mark off the behavior of the person you'll be seeing into the largest units that seem natural and meaningful to you." This procedure was modeled after that of Newtson (Newtson, 1973), with the addition of the verbal protocol. We will refer to these as "fine" and "coarse" coding conditions, respectively.

Participants first segmented the example tape, and then each of the four activities. Each participant saw two activities performed by each actor. The order of activities, actors, and the pairing of actors to tapes was varied for each subject to minimize order effects (but not fully counterbalanced, as that would have required 96 participants).

After viewing all four activities, participants engaged in an unrelated experiment for about 25 minutes. Then, they watched the same four tapes in the same order. This time however, they were given the opposite unit-size instructions: if they had been instructed to use the "smallest" units (fine coding condition) before, now they were told to use the "largest" units (coarse coding), and vice versa.

Verbal responses were recorded with a cassette recorder. Tapping times were recorded by a Macintosh IIci computer connected to the keyboard, running a simple script written in PsyScope 1.1 (Cohen, MacWhinney, Flatt, & Provost, 1993).

Event Segmentation Analyses

Discrete Analysis

First, the tapping record for each participant viewing each videotape was divided into 1-second bins. All the results reported here are based on 1-second bins, but to the extent we have been able to verify, they hold across bin sizes from 1 to 5 seconds. Following Newtson's (1973) terminology, each bin was coded as a "breakpoint" if it contained one or more taps. Bins which were breakpoints for a given subject in both the fine and coarse coding conditions were called "overlaps." For each participant and each tape they saw, the following were calculated:

Bins = number of bins in the tape

Fine = number of breakpoints in the fine coding condition

Coarse = number of breakpoints in the coarse coding condition

P(fine) = probability that a given bin will be a fine breakpoint = Fine/Bins

P(coarse) = probability that a given bin will be a coarse breakpoint = Coarse/Bins

Overlaps = number of bins that were breakpoints in both the fine and coarse coding conditions

Now, suppose there is no relationship between coarse and fine unit boundaries (i.e. they are independent). Under this assumption, the probability that a given point in a videotape will be identified as a breakpoint in the fine coding condition is independent of the probability that it will be identified as a breakpoint in the coarse coding condition. Under this assumption, the expected number of overlaps can be approximated as:

$$(1) \quad \text{Overlaps}_0 = \text{P(coarse)} \times \text{P(fine)} \times \text{Bins}$$

Equation (1) can be easily calculated by expanding to:

$$\text{Overlaps}_0 = \frac{\text{Fine}}{\text{Bins}} \times \frac{\text{Coarse}}{\text{Bins}} \times \text{Bins} = \frac{\text{Fine} \times \text{Coarse}}{\text{Bins}} \quad (2)$$

Equations (1) and (2) give us a null model that can be compared to the actual number of overlaps. This is essentially a within-subject version of the analysis reported by Newtson (1973).

Continuous Analysis

The discrete analysis is attractive because its statistical properties are easily understood, but it has the disadvantage of depending on an arbitrary choice of a discrete bin size. As an alternative, we also developed a continuous analog of the discrete analysis. As with the discrete analysis, this approach compares the two viewings of each tape for each participant. Here, "breakpoint" refers to the actual time of a tap. Breakpoints in the fine coding condition will be called "fine breakpoints," and breakpoints in the coarse coding condition "coarse breakpoints." For each coarse breakpoint, the distance to the nearest fine breakpoint was calculated. These distances were averaged across the coarse breakpoints for a given participant watching a given tape to calculate:

AvgDist = mean distance from coarse breakpoints to the nearest fine breakpoint for a given pair of viewings of an activity by a given participant

Now, as in the discrete case above, a null model is required to which to compare these scores. In this case, one can calculate an expectation for AvgDist given independence of the coarse and fine breakpoints. Begin by taking the location of the fine breakpoints as given. Generate coarse breakpoints distributed randomly and uniformly across the tape, and measure their distance to the nearest fine breakpoint. In the limit case, this amounts to integrating the distance to the nearest fine breakpoint over the length of the tape. If the location

of the last fine breakpoint ($f_n$) is taken as an estimate of the length of the action on the tape for that viewer in milliseconds and F= $\{f_1,f_2,...,f_n\}$ (where $f_1,f_2,...,f_{Fine}$ is the set of all fine breakpoints of this participant while watching this tape, in milliseconds), then the null prediction is :

$$\text{AvgDist}_0 = \frac{\frac{f_1^2}{2} + \sum_{i=1}^{i=n-1}\left[\frac{f_{i+1}-f_i}{2}\right]^2}{f_n} \quad (3)$$

(Note that the first term in the numerator is just the special case at the beginning of the tape.)

The two analytic methods are illustrated in Figure 2.

---

Insert Figure 2 about here

---

Results

All results reported here are based on the Describe group, except the comparison of the Describe and Silent groups.

There were a few cases in which participants denoted very long units, which fell outside the distribution for the experimental group (13 units whose length was greater than two standard deviations from the mean, all of which occurred in the coarse viewing condition and 11 of which occurred for familiar tapes). These corresponded to viewings on which the participant tapped only once or twice. Because the analyses reported here could be sensitive to the influence of a small number of outlying observations, data from these 13 viewings were removed from further analysis (except as noted). There were also 4 viewings during which the computer recorded no taps at all; these were also excluded.

Participants easily segmented the activities at either a fine or coarse grain. The length of segments produced under the fine-unit coding instructions was substantially shorter than that produced under the coarse-unit coding instructions. Overall, for the Describe group the mean length of coarse-unit breakpoints was 34300 ms(SEM = 2610 ms), and the mean length of fine-unit breakpoints was 12800 ms (SEM = 1040 ms). This corresponds to a mean of 10.1 breakpoints per coarse-unit viewing (SEM = 0.92), and 28.9 breakpoints per fine-unit viewing (SEM = 2.02). For both conditions, there were reliable differences in mean unit length between the four activities. These were assessed with separate analyses of variance blocked on participant for the coarse and fine conditions: For the coarse-unit coding condition, $F(3,77) = 2.87$, $p = .04$; for the fine-unit coding condition, $F(3,92) = 17.9$, $p < .001$. (The differing degrees of freedom reflect small differences in the number observations.)

---

Insert Figure 3 about here

---

However, there were also considerable individual differences between participants in the rate of segmentation in the fine and coarse conditions, which can be seen in the overlap between the distributions in Figure 3. Given the robust individual differences in natural segmentation level, aggregating breakpoints across individuals presents something of a challenge. Nonetheless, there was modest agreement as to the location of coarse and fine breakpoints, as can be seen in Figure 4. Moreover, these individual differences recommend the use of within-participants evaluations of alignment between coarse and fine breakpoints, as were performed here.

The ratio of fine-unit breakpoints to coarse-unit breakpoints was somewhat stable across individuals. The median ratio was 3.15, and for 24 of the 32 participants it was between 1 and 5. It is striking that the modal pattern of decomposition across temporal grains was to break each coarse unit into roughly three fine units. One possibility is that the schema "beginning, middle, end" has perceptual priority.

---

Insert Figure 4 about here

---

**Presence of Hierarchical Structure**

The segmentation data for the Describe group were analyzed using both the discrete and continuous methods. The first question asked was: Do the coarse and fine break points fall into alignment more than chance predicts? For the discrete method, we calculated Overlaps and $Overlaps_0$ for each participant's viewing of each tape from the fine and coarse codings. On average there were reliably more overlaps per viewing (Overlaps mean = 2.57, SEM .329) than was predicted by the null model ($Overlaps_0$ mean = 2.00, SEM .284), $t(146) = 4.42$, $p < .001$. This provides clear evidence for hierarchical structure. Results from the continuous method were consistent with those from the discrete method. For each participant's viewing of each tape, we calculated AvgDist and $AvgDist_0$ from the fine and coarse codings. Overall, the mean distance per viewing from each coarse breakpoint to the nearest fine breakpoint was on average closer (AvgDist mean = 2380 ms, SEM = 303) than predicted by the null model ($AvgDist_0$ mean = 4410 ms, SEM = 388), $t(110) = 8.64$, $p < .001$, again supporting the hypothesis of hierarchical structure. Thus, both analyses indicate the

presence of an alignment effect: unit boundaries under the coarse and fine coding conditions were in better alignment than would be predicted by chance.

One might be concerned that these results reflect participants' memory during the second viewing for the locations at which they segmented the activity during the first viewing. To address this concern, we adapted the continuous analysis to compare first-viewing data between participants[1]. For each participant whose first viewings were under coarse coding instructions, we compared the location of their coarse breakpoints to the fine breakpoints of the participants who had seen the same tape for the first time under fine coding instructions. The results were consistent with the previous analysis. For the first-viewing data, the mean $AvgDist_0$ was 4670 ms (SEM = 638), while the mean AvgDist was 2820 ms (SEM = 151), $t(54) = 2.78$, $p = .0075$. As expected, the size of the effect is smaller and the variability of the difference between $AvgDist_0$ and AvgDist is larger (a standard deviation of 4920 ms for the between-participants analysis, compared to 2480 ms for the within-participants analysis), reflecting the fact that participants did not always agree on breakpoint locations. Moreover, we urge some caution in interpreting this analysis, given the large individual differences in overall coding level. That being said, the fact that this analysis was able to detect an alignment effect in the presence of those individual differences is a further indication of the robustness of the alignment effect.

**Effects of Familiarity**

We investigated the effects of familiarity on both segmentation level and degree of alignment between coarse and fine breakpoints. To test effects of

---

[1] We would like to thank Yaakov Kareev for suggesting this analysis.

familiarity on segmentation level, we calculated the mean unit length for each participant's observation of each tape, for all subjects in the Describe group. The scores were then submitted to an ANOVA with familiarity and condition as factors, blocked on subjects. There was no main effect of familiarity on unit length, $F(1,204) = .440$, $p = .508$, and no interaction with condition, $F(1,204) = 1.51$, $p = .220$. Thus, familiarity did not reliably affect the length of perceived units. (A note regarding the outliers: Transforming the untrimmed scores with a log transformation gave the same results, while analyzing the untrimmed scores led to both a main effect of familiarity and an interaction between condition and familiarity, as would be expected from the location of the outliers.)

To test effects of familiarity on degree of alignment, we conducted analyses based on the discrete and continuous methods described above. First, we calculated for each participant's viewing of each tape a difference between the observed number of overlaps and the number predicted by chance (Overlaps - $Overlaps_0$). These scores were submitted to an ANOVA with familiarity as the only factor, blocked on subjects. This difference was larger for familiar activities (mean = .857, SEM = .165) than for unfamiliar activities (mean = .327, SEM = .159), and this difference was statistically reliable, $F(1,82) = 5.37$, $p = .02$. Second, we calculated for each participant's viewing of each tape the difference between the mean distance from a coarse breakpoint to the nearest fine breakpoint and that expected by chance. These scores were also submitted to an ANOVA with familiarity as the only factor, blocked on subjects. The pattern was consistent with that obtained from the discrete analysis: Mean distances were on average closer than predicted by chance by a greater degree for the familiar activities

(mean = 2.19 s, SEM = .375) than for the unfamiliar ones (mean = 1.90 s, SEM =. 297). However, this effect was not statistically reliable, $F(1,78) = 2.278$, $p = .44$.

To follow up these suggestive results, we conducted a further analysis. For each of the individual coarse breakpoints, the distance to the nearest fine breakpoint in the same observer's viewing of the same videotape was computed, as in the continuous analysis above. However, instead of averaging these distances and comparing them to null model, we submitted the distances themselves to an ANOVA with familiarity as a factor and videotape as a factor nested on familiarity, blocked on subject. On average, coarse breakpoints were closer to their nearest fine breakpoint for familiar activities (mean = 1470 ms, SEM = 73.6), than for unfamiliar ones (mean = 1820, SEM = 136), and this was highly reliable, $F(1,1087) = 7.48$, $p = .006$. (There was also an effect of the nested variable videotape, $F(1,5) = 5.48$, $p < .001$, indicating that familiarity did not account for all the differences between the videotapes in degree of alignment.) Several comments about this analysis are in order. It has the advantage of achieving greater power by analyzing the individual distances rather than means per viewing, but has the disadvantage of not allowing a comparison with the null model expectation for each pair of viewings. It is possible that familiar and unfamiliar activities differed on some extraneous feature that caused their distance scores to vary but would also have affected the expected distance scores, if they were available. Also, it should be noted that this analysis weights the contributions of participants who made finer units (and thus contributed more data) relative to the other analyses. In spite of these reservations, the converging evidence from all three analyses suggests that there was indeed

greater alignment between coarse and fine breakpoints for the familiar activities than for the unfamiliar ones.

**Effects of Describing**

How does verbally describing activity as it happens affect perception of that activity? In particular, does adding verbal description to the segmentation task affect the alignment of coarse and fine breakpoints? There are two obvious, and conflicting, predictions. The addition of verbal description to the segmentation task yields a dual task design. To the extent that the two tasks share common processing resources, performing either task should interfere with performance of the other. By an attentional account, then, interference on the segmentation task should add noise to the tap locations, resulting in a lower degree of alignment between coarse and fine breakpoints when describing activity. On the other hand, if people are disposed to encode activity in terms of hierarchical schemata, and if these schemata are constituted in part as propositional or quasi-verbal representations, describing activity as it happens may increase the influence of these representations, leading to an increase in the alignment effect. To examine this question, we applied the continuous and discrete analyses to a comparison of the segmentation data from the Describe and Silent groups.

To investigate the influence of the verbal description task on the alignment effect, we first applied the discrete analysis. For each viewing, the number of overlaps (Overlaps) and the number of expected overlaps ($Overlaps_0$) were calculated, and a difference score obtained. These scores were submitted to between-groups ANOVA with subject as a nested factor within group. The mean difference score was larger for the Describe group (mean = .571, SEM =

.117) than for the Silent group (mean = .429, SEM = .377), though this difference was small and statistically unreliable, $F(1,107) = .216$, $p = 0.64$. We also applied the continuous analysis. For each viewing, the mean distance from each coarse breakpoint to the nearest fine breakpoint (AvgDist) and its expectation ($AvgDist_0$) were calculated and difference scores obtained. These scores were submitted to an ANOVA as was done for the discrete scores. Consistent with the discrete analysis, the alignment effect was larger by this analysis for the Describe group (mean = 2030 ms, SEM =235) than for the Silent group (mean = 1340 ms, SEM = 320), and this difference was marginally statistically reliable, $F(1,103) = 3.26$, $p = .07$. To follow up these results with a more powerful analysis, we analyzed the raw distance scores as was done for familiarity. For each of the individual coarse breakpoints, the distance to the nearest fine breakpoint in the same observer's viewing of the same videotape was computed, as in the continuous analysis above. However, instead of averaging these distances and comparing them to null model, we submitted the distances themselves to an ANOVA with group as a factor, and subject nested within group. The distance from coarse breakpoints to their nearest fine breakpoint was on average smaller when describing while segmenting (mean = 1620 ms, SEM = 72.0), than when not (mean = 3680 ms, SEM = 251), and this difference was highly reliable, $F(1,1397) = 147$, $p < .001$. The caveat that applied in the use of this analysis for the familiarity comparison does not apply here, as there is no comparison across items. The comment that this analysis disproportionately weights the contributions of participants who tapped more frequently still applies.

We also analyzed the effect of verbal description on unit size. The mean unit length for each viewing of each tape was calculated for all subjects in both

the Describe and Silent groups. Outliers were eliminated using the same cutoff as in the analysis of familiarity described previously. (There were no outliers in the segment-only group.) The data were analyzed with a between-groups ANOVA, with subject nested on group. Participants divided the activity into slightly larger units when asked to describe the activity (34.3 s vs. 28.5 s for coarse, 12.8 s vs. 10.7 s for fine). However, this pattern was not statistically reliable, $F(1,261) = 2.67 = .10$; neither was the interaction between group and coding level, $F(1,261) = 0.911$, $p = .34$.

On-line Descriptions of Events

From audiotapes of the 32 participants in the Describe group, 16 were selected for transcription and analysis (based on audibility of the recording). Eight had been run with the fine unit coding instructions given first, and 8 with the fine unit instructions given second.

Audiotapes were transcribed and then coded by two judges. Each transcribed utterance was recorded along with the location of the key tap that marked the end of the unit it described. Utterances that consisted of two sentences, or two independent clauses joined by a conjunction were recorded separately, with the same unit index. Each utterance was rated on several features by both coders. These features described the subject, verb, and up to 3 objects (direct objects, indirect objects, or objects of prepositions) per utterance. (Of the 3171 utterances coded, 2 had 4 objects. For these two, the 4th object was left off.)

Characteristics of the Descriptions

A few words are in order regarding the coding. The vast majority of utterances described actions on objects. (94.5% contained a verb and at least one

direct or indirect object or object of a preposition.) The major exceptions were initial and final segment descriptions, which often described the actor entering/exiting the room. Given that only a single actor was involved in each activity, subject reference was not expected to be revealing and was in fact dropped in most utterances. The measures we examine reflect primarily presupposition and generality of reference to objects and actions. Presupposition is a good clue to horizontal segmentation. Within a segment, the same elements are relevant, so they can be presupposed. Defining elements might more likely to be explicitly mentioned at segment boundaries. Ellipsis, pronominalization, and marking of recurrence are all signs of presupposition. We use the word "recurrence" to refer to the linguistic marking of a subject, verb or object as a member of a set or group. For example, in the utterance "tucking it <u>again</u>," the verb is recurrent; in the utterance "puts on <u>second</u> corner" the object is recurrent. Subjects and objects could be subject to pronominalization or ellipsis, but not both (because one cannot pronominalize a subject or object that is not said). Generality is an indexing focus. When objects are more focal, they are more likely to be referred to specifically.

<u>Subject Pronominalization/Ellipsis</u>: Was the subject of the sentence pronominalized or left off (elided) (P, E, or neither)?

<u>Subject Recurrence</u>: Was the subject marked as recurrent (T/F)? (Subject recurrence never occurred.)

<u>Verb Ellipsis</u>: Was the verb left off (elided) (T/F)?

<u>Verb Recurrence</u>: Was the verb marked as recurrent (T/F)?

For each object, the following were coded:

Object Pronominalization/Ellipsis (per object): Was the object pronominalized or elided (P, E, or neither)?

Object Recurrence (per object): Was the object marked as recurrent (T/F)?

The two coders worked independently, and disagreements were adjudicated by the first author. Based on the object ratings, the following composite scores were computed for each utterance:

Object Ellipsis: Proportion of objects in the utterance subject to ellipsis (0-1).

Object Pronominalization: Proportion of objects subject to pronominalization (0-1).

Object Recurrence: Proportion of objects marked as recurrent (0-1).

In addition, several other features were coded automatically or semi-automatically:

Progressive: Was the verb in the progressive, as opposed to the perfect, form (T/F)?

Number of objects in the utterance (1-3).

Verb Repetition: Was the verb repeated from the previous utterance to this one (T/F)?

Object Repetition: Were any of the objects repeated from the previous utterance to this one (T/F)?

From the WordNet database (Fellbaum, 1998, version 1.6), we obtained polysemy measures for the objects and verbs. The number of senses for verbs appearing in the transcripts ranged from 0 (for verbs that did not appear in the lexicon) to 48, and the number of senses for objects ranged from 0 to 19. Based on these measures, the following were calculated:

Verb Polysemy: The number of senses in the WordNet database for the verb in the utterance (0-48).

Object Polysemy: The mean number of senses in the WordNet database for the objects in the utterance (0-19).

Finally, we obtained ratings for each of the verbs for goal-directedness and generality, and each of the objects for generality (see Appendix C). From these ratings the following were calculated:

Verb Generality: How general was the verb (1 = "very specific" to 5 = "very general")?

Verb Goal-directedness: How goal-directed was the verb (1 = "very goal-directed" to 5 = "non goal-directed").

Object Generality: The mean generality rating (How general was each object?, 1 = "very specific" to 5 = "very general").

These ratings were the input to the analysis of the descriptions.

Each utterance corresponded to a fine or coarse unit. To examine the relationship between segmentation structure and the content of the verbal descriptions, we subdivided the fine unit descriptions into two classes. Boundary units were fine units that corresponded to the boundaries between coarse units. Boundary units were identified by finding, for each coarse unit, the nearest fine unit breakpoint. This was taken to be the end of the terminal fine unit in that coarse unit. The following fine unit was taken to be the initial fine unit in the next coarse unit. Hence, both were marked as boundary units. All the fine units not so marked will be called internal units.

For each participant's two viewings of each activity, all the features described above were tabulated, broken down as coarse-, boundary-, or

internal-unit descriptions. For T/F features, a proportion was computed, and for numerically coded features a mean was computed.

The linguistic analysis addresses several issues. First, is the structure that is evident in the segmentation data also present in the verbal descriptions of activity? In other words, Does the hierarchical bias hypothesis hold for descriptions of activity as well as for temporal segmentation of activity? This question can be sharpened by considering the distinction between boundary and internal fine units. Under the hypothesis that boundary units correspond to cognitive coarse-unit boundaries as well as fine-unit boundaries, descriptions of boundary units should be more similar to coarse units than the rest of the fine units. We evaluated this hypothesis with regard to all the features tested. A second objective of the analysis was to characterize the salient features of event segments at both coarse and fine levels. This should give insight into the internal structure of event segments and may reveal qualitative differences between the levels. Thus, the descriptions inform us as to how events are thought about both vertically (across segmentation levels) and horizontally (across time within a segmentation level).

Because a large number of features were explored, no hypothesis tests will be reported. Rather, we report means of the scores with 95% confidence intervals.

Results

Insert Figure 5 and Figure 6 about here

Both fine- and coarse-unit utterances were by and large telegraphic,

concrete descriptions of individual actions on objects. Figure 5 and Figure 6, which present transcripts of two sets of transcripts, illustrate the typical pattern. The 15 objects that occurred most often in the corpus were (in order of frequency): plate, sheet, pillow, saxophone, bed, apron, drawer, water, dishwasher, pillowcase, something, glass, silverware, plant, and box. The 15 verbs that occurred most often were (again in order of frequency): put, take, pick, open, close, wash, turn, rinse, tuck, walk, leave, place, pull, scrape, water, pour. As can be seen from these verbs and the transcripts, utterances conveyed basic intentional acts.

Two general phenomena are evident in the linguistic analysis. We will provide qualitative characterizations of each before presenting the quantitative details. First, for most of those differences between coarse-unit and fine-unit descriptions, the boundary units had values intermediate between those of the coarse units and the rest of the fine units, as predicted by the hierarchical bias hypothesis.

Second, descriptions of coarse and fine units differed systematically. In general, coarse-unit descriptions specified objects more precisely than fine-unit descriptions. By contrast, fine-unit descriptions specified verbs more precisely than coarse-unit descriptions (with two exceptions, as noted in the following). Qualitatively, in the coarse-unit descriptions participants seemed to be carefully identifying some set of objects or part of the environment with the coarse-unit descriptions—locating the action—but characterizing just what was happening less clearly. In the fine-unit descriptions, they were less careful in describing the objects involved but precisely specified what was happening.

To examine these phenomena in detail, we turn first to object features. These data are presented in Figure 7. In general, participants tended to mention objects by name, but they did use several syntactic short cuts on occasion. They sometimes elided objects or referred to them with pronouns, they marked them as recurrent, and then often repeated the identical object from utterance to utterance. All four of these syntactic devices occurred more often in coarse-unit than fine-unit descriptions, with boundary units in between. Together, these results indicate that participants employed short cuts for objects more often when describing internal fine units than when describing coarse units, with boundary fine units falling in between. Presumably objects can be referred to more vaguely for fine units because there is less ambiguity in the referent.

---

Insert Figure 7 about here

---

The semantics of the object descriptions shows the same pattern, though less strongly. Participants described objects with more polysemous words under fine coding conditions than under coarse coding conditions, with boundary units in between. However, the ratings of object generality show a weak trend in the opposite direction: Objects were rated slightly more general in the coarse-unit descriptions than in the two groups of fine-unit descriptions.

The modal number of objects per utterance was one. There were marginally more objects used in fine-unit descriptions than in coarse-unit descriptions.

Now, we turn to the verb features, presented in Figure 8. Verbs were rarely elided or marked as recurrent, but when this did happen it was more likely to occur under fine-unit coding conditions than under coarse-unit

conditions, with boundary-fine-units in between the internal-fine-unit and coarse-unit descriptions for both. These weak patterns run counter to the general characterization that verbs were characterized more vaguely in the coarse-unit coding condition.

There was a larger effect in the pattern of verb repetition: verbs were more likely to be repeated under coarse-unit coding conditions than under fine-unit coding conditions. This pattern does accord with the general characterization.

Use of the progressive aspect did not differ appreciably across coding conditions. Verb aspect in this task appeared to be an individual difference variable: A given observer tended to choose a single aspect and stick with it throughout the experiment.

---

Insert Figure 8 about here

---

Semantically, participants tended to use more polysemous verbs when describing coarse units or boundary fine units than when describing internal fine units. They used verbs that were rated more general when describing coarse units than internal-fine units, with boundary units falling in between. Verbs from coarse-unit descriptions were also rated more goal-directed than verbs for internal-fine units, with boundary units again falling in between.

Finally, we turn to the features pertaining to aspects of how subjects were described (see Figure 9). The subject of the descriptions was almost always the actor in the videotape. It is therefore not surprising that the subject was elided or referred to by pronoun, and never marked as recurrent. Pronominalization occurred more often for coarse units than for internal-fine units; ellipsis occurred

more often for internal-fine units than for coarse units. For both features, boundary fine units fell in between the coarse units and the internal fine units.

---

Insert Figure 9 about here

---

To summarize, of the 17 syntactic and semantic features that varied in the descriptions, only 2 violated the pattern of boundary-fine units taking a value intermediate between those of coarse and fine units. In general, objects were specified more precisely in coarse units than in fine units while verbs were specified more precisely in fine units than in coarse units, with the exception of object generality, verb ellipsis and verb recurrence.

Discussion

Participants twice watched videotapes of two familiar events, making a bed and doing the dishes, and two unfamiliar events, assembling a saxophone and fertilizing houseplants. On one viewing, they segmented the events into the largest units that seemed natural and meaningful; on the other viewing, they segmented the events into the smallest units that seemed natural and meaningful. Some of the participants described the units as they segmented. The major question is whether the events are perceived hierarchically. Evidence supporting hierarchical organization comes from both the segmentation and the descriptions. In addition, the descriptors provide insight into the peoples' conceptions of events and their temporal structure.

Segmentation of continuous activity

The segmentation data from this experiment showed three distinctive patterns. First, across experimental manipulations the locations of unit boundaries under fine and coarse coding conditions were in closer alignment

than was predicted by an appropriate null model, the alignment effect. This was verified by two converging analytic strategies, the first based on discretizing the time-line and counting overlaps between tap locations under coarse and fine coding conditions, the second based on the continuous locations of the perceptual unit boundaries. It was reliably observed on both a within-participants and between-participants basis. The alignment effect constitutes clear support for the hierarchical bias hypothesis.

Second, despite the dual-task demands on participants, the alignment effect was more pronounced when participants described the activities while segmenting them than when they only segmented. This was borne out by both the continuous and discrete analyses. This suggests that in order to talk about activity coherently at a single temporal grain, observers spontaneously draw on mental representations of the activity that contain information about relationships across temporal grains.

Third, the alignment effect was slightly more pronounced for familiar activities than for unfamiliar activities. A series of three analyses converged on this conclusion. There was little evidence that observers under these conditions segmented familiar activity into larger units than unfamiliar activity.

Together, these features of the on-line segmentation data support the hypothesis that observers are disposed to encode activity in terms of units organized as a partonomic hierarchy.

Describing Events

The alignment of coarse and fine unit boundaries provides strong evidence that perceivers' understanding of unfolding events is based on partonomic hierarchies. The simultaneous descriptions of coarse and fine unit

activity not only corroborate the psychological reality of a hierarchical knowledge structure, but also help to characterize that knowledge.

On the whole, the descriptions were brief, telegraphic. They lacked the sometimes chaotic form of conversation and even lacked the communicative intent of a radio sports announcer (e.g, Clark, 1996). The vast majority of descriptions were of the form: action on object. In Talmy's analysis, a motion event consists of a figure, a motion, a path, and a ground (Talmy, 1975). For most cases, these telegraphic descriptions omitted both figure and ground, presumably because they were constant throughout each film and could be presupposed. The descriptions did include the motion in the verbs and paths in the verb particles. These descriptions of actions on objects expressed functional, causal, goal-oriented, purposeful relations. This need not have been the case. The descriptions at one or both levels could have referred to activities of the body, such as raising the arms, clenching the fists, or bending the waist, or even to states, such as standing or leaning. Instead, the descriptions referred to accomplishments or achievements, activities that culminate in natural endings (see Casati & Varzi, 1996). The descriptions, then, strongly suggest that perception of unfolding events entails thinking about function, causes, goals, and ends.

The organization of event descriptions closely paralleled the behavioral segmentation data. On almost every semantic and syntactic measures of subjects, objects and verbs, the boundary fine units fell in between coarse and fine units. Even when observers were segmenting at a fine level those portions of activities that turned out to align with coarse unit boundaries were perceived as special, as different in status than the other fine units. In other words, the

hierarchical structure observed in the segmentation data was replicated within the fine-unit descriptions. This was true despite the facts that (a) participants were unaware of the distinction between boundary and internal units, (b) the experiment instructions were very specific in not asking participants to provide any information about the relationships among units, and (c) the features of the experimental situation did not encourage a conversational mode of speech.

Although both coarse and fine units evoked descriptions of actions, coarse and fine units evoked qualitatively different descriptions. According to a number of measures, both syntactic and semantic, descriptions of coarse units referred to objects precisely but to actions vaguely. Conversely, descriptions of fine units tended toward vaguer references to objects and more precise references to actions. Put differently, different coarse units differed from one another by the object of interaction, and by implication, by action as different objects often require different actions. In contrast, different fine units belonging to the same coarse unit differed from one another on the action performed on the same object. The boundary fine units fell between coarse and fine units on most measures.

The qualitative differences in descriptions at coarse and fine levels of event segmentation support an object/action account of event structure perception. We briefly outline the account here, and will return to a fuller explication in the General Discussion. The finding that references to objects were vaguer at the fine level than the coarse level suggests that coarse units tend to be punctuated by objects. Fine units within the same coarse unit presuppose the same object. Put differently, the same object is focal for the entire coarse segment. This is substantiated by the finding that references to actions were more specific at the

fine level than the coarse level; that is, different fine units tended to differ on actions. Within the fine units belonging to the same coarse unit, then, refined actions on the same object are focal. Not only is the segmentation of events hierarchical, but there are qualitative differences in the levels of the hierarchy.

One final finding deserves attention. Describing events while segmenting them yielded a greater degree of alignment between coarse and fine units. This suggests that segmentation is determined by both bottom-up perceptual differences in activity and top-down knowledge about event structure. Actively describing the contents of each segment appears to invoke top-down knowledge structures, greater awareness of causal, functional, and intentional relations. This in turns suggests that using language, and perhaps language itself, biases away from raw perceptual statements and toward causal and intentional ones.

## Experiments 2 and 3: Manipulations of Familiarity

The alignment effect is the most striking result of the perceptual analyses in Experiment 1: coarse and fine event segment boundaries aligned more than would be predicted by chance. The fact that it was influenced by the familiarity of the activities supports the view that event structure perception is mediated by hierarchically organized schemata. When the knowledge structure was more developed, coarse and fine-unit segmentation was more aligned. The hierarchical bias hypothesis suggests further relations between familiarity and alignment.

First, increasing the familiarity of the activity to be segmented should increase the alignment effect. One way to make an unfamiliar activity more familiar is by teaching it. It has been argued that the crux of teaching a complex

procedure is providing the learner with an appropriate structured mental model (Kieras, 1988). According to the hierarchical bias hypothesis, an appropriate model should provide the learner with an appropriate partonomic decomposition of the activity. This top-down structured knowledge should increase the alignment effect in the perceptual segmentation paradigm. In Experiment 2, participants were taught an unfamiliar activity, assembling a saxophone, in the laboratory. It was hypothesized that this training would increase the magnitude of their alignment effect.

Second, populations with greater a priori familiarity with an activity should show a greater alignment effect. In particular, for a given activity experts should show more of an alignment effect than novices. In Experiment 3, experts and novices at saxophone assembly were directly compared. It was hypothesized that the experts would show a greater alignment effect for the videotape of assembling a saxophone, but not for the other videotape.

Another motivation for these studies was a concern that the familiarity effect observed in Experiment 1 might be due to particulars of the small number of familiar and unfamiliar items sampled. Both the teaching manipulation and the expert-novice comparison address this issue.

These experiments take two complementary approaches to understanding the effects of familiarity on event perception. They also provide replications of the alignment effect found in the first experiment. As will be seen, the alignment effect replicated vigorously. However, increasing familiarity did not increase the alignment effect in either experiment. Experiment 3 indicated that the failure of the familiarity manipulations may be due to weakness of the familiarity effect itself, for the materials used here.

Method

The methods employed were almost identical to that of Experiment 1. Differences will be noted below.

Materials

For Experiment 2, one stimulus was selected from the eight employed in Experiment 1: the videotape in which the female actor assembled a saxophone. It was chosen because it was the least familiar of the four activities studied in Experiment 1. For Experiment 3, videotapes of all four activities performed by the female actor were used. In both experiments, the same example stimulus was used (in which a woman ironed a shirt).

Participants and Procedure

In Experiment 2, participants were screened to be unfamiliar with the saxophone, and randomly assigned to two different groups. One group (the trained group) was given a short course in putting together a saxophone before beginning the event-segmentation phase of the experiment. In this training, the experimenter demonstrated how to put together a saxophone and identified all the parts of the instrument. The training was complete when the subjects were able to accurately recall all the names of the different parts of the saxophone twice successively. The training procedure lasted about eight minutes. The other group (the untrained group) received no training. Twelve participants were randomly assigned to each group.

In Experiment 3, expert saxophone-assemblers were recruited from the Stanford Band and compared to novices. To minimize possible confounding

variables, novices were also selected from a local musical ensemble: violinists from the Stanford Symphony. Sixteen participants were selected for each group.

Participants received $8 or course credit in Psychology for participating.

The rest of the procedure in both experiments was essentially the same as in Experiment 1. The instructions (including the coarse/fine manipulation) were identical. Participants received fine or coarse coding instructions, then first segmented the example tape, then the experimental tape(s). They then performed an unrelated task for 25 minutes, after which they segmented the experimental tape(s) again, under the alternative coding instructions. Order of coding instructions was counterbalanced within each group, and order of stimulus presentation in Experiment 3 was controlled as in Experiment 1.

One feature of the procedure in Experiment 3 differed between from the previous ones. For this experiment, the video stimuli were presented on a computer (an Apple Macintosh equipped with an Avid Cinema video digitization/compression card), using the same PsyScope (Cohen et al., 1993) script that was used to collect the tapping data. This allowed for more precise timing of the breakpoint locations, and automatic synchronization between the video stimulus and the tapping data.

Results

The alignment effect of Experiment 1 was replicated in both experiments: Coarse breakpoints were on average closer to the nearest fine breakpoint than was predicted by the appropriate chance model. This was demonstrated by both the discrete and continuous analysis methods (see Experiment 1).

For the discrete method, there were reliably more overlaps per viewing than was predicted by the null model. In Experiment 2, the difference was .874,

$t(23) = 3.4$, $p = .002$. In Experiment 3, the difference was 1.52, $t(127) = 7.1$, $p < .001$. For the continuous method, the mean distance from each coarse breakpoint was on average closer than predicted by the null model. In Experiment 2, the difference was 3420 ms, $t(23) = 7.04$, $p < .001$. In Experiment 3, the difference was 3450 ms, $t(127) = 12.1$, $p < .001$.

The influence of training and expertise on the alignment effect was tested with both the discrete and continuous analytic methods. Expected and observed scores were calculated, and the difference between the two was submitted to a between-participants ANOVA. The results were inconclusive. In Experiment 2, the discrete analysis indicated an effect of training on alignment, opposite to that predicted. The difference between the observed number of overlaps and that predicted by chance was smaller for the trained group (mean = .359, SEM = .232) than for the untrained group (mean = 1.39, SEM = .418), and this difference was statistically reliable, $F(1,22) = 4.68$, $p = .04$. However, the mean distance per viewing from coarse breakpoints to nearest fine breakpoint did not differ appreciably between the trained group (mean = 3400 ms, SEM = 902) and the untrained group (mean = 3430 ms, SEM = 412), $F(1,22) = 0.00$, $p = .97$. The results are summarized in Figure 10.

---

Insert Figure 10 about here

---

In Experiment 3, the expert group was predicted to show an especially large alignment effect for the "assembling a saxophone" activity, but not for the other three. This interaction between expertise and activity was not observed, as can be seen in both panels of Figure 11. This was borne out by both the discrete analysis, $F(3,90) = .490$, $p = .69$, and the continuous analysis, $F(3,90) = .771$, $p = .51$.

The discrete analysis showed no indication of a main effect of group, $F(1,90) = .012$, $p = .91$; however by the continuous analysis the novice group showed a larger alignment effect overall, $F(1,90) = 5.28$, $p = .02$.

Figure 11 also indicates that the basic familiarity effect replicated only weakly in this experiment. By both analyses there was a trend in the direction of replicating the original familiarity effect, but it was not reliable for the discrete analysis, $F(1,124) = 1.87$, $p = .17$, or for the continuous analysis, $F(1,124) = 3.43$, $p = .07$.

---

Insert Figure 11 about here

---

Neither training nor expertise affected the level at which participants segmented the activity. Mean unit lengths per viewing were calculated and submitted to ANOVAs with training as a between-subjects factor, condition as a within-subjects factor, and subject blocked on training. In Experiment 2, there was no effect of training, $F(1,22) = .023$, $p = .88$, and no interaction with condition, $F(1,22) = .506$, $p = .48$. The expected difference in unit length between the coarse and fine coding conditions was highly reliable, $F(1,22) = 23.2$, $p < .001$. In Experiment 3, there was no effect of expertise, $F(1,254) = .576$, $p = .45$, and no interaction with condition, $F(1,254) = .413$, $p = .521$. In both cases the differences between coarse and fine unit lengths were large and reliable, as expected.

Discussion

These experiments replicated the alignment effect of Experiment 1, further supporting the hierarchical bias hypothesis. However, neither training of novices (Experiment 2) nor direct comparison of experts and novices (Experiment 3) mediated the familiarity differences.

While the groups did not differ in the degree to which they hierarchically organized the activity in either Experiment 2 or Experiment 3, they might differ qualitatively in where they segmented the behavior. To investigate this possibility we plotted probability densities of breakpoint locations for each of the groups in each of the coding conditions. In both experiments, the distributions were quite similar between the two groups, suggesting they did not perceive qualitatively different segmentations. We also performed a statistical test of the amount of disagreement across the two groups for each coding condition, following the method described by Massad and his colleagues (Massad, Michael, & Newtson, 1979). Briefly, this analysis identifies the most common breakpoints in each group, and then tests to see whether the proportion of observers who segmented the activity at each of these points differs across groups[2]. We performed this analysis for both Experiments 2 and 3, independently treating the coarse and fine coding conditions. There was no evidence that segmentation patterns differed between trained and untrained novices, or between experts and novices. (In Experiment 2, for the fine-unit breakpoints $F(1,40) = 0.211$, $p = .65$, for the coarse-unit breakpoints $F(1,24) = 0.480$, $p = .50$. In Experiment 3, for the fine-unit breakpoints $F(1,40) = 0.244$, $p = .623$; for the coarse-unit breakpoints $F(1,8) = 0.15$, $p = .71$.) While it would be difficult to observe a reliable difference in the coarse coding conditions due to the small number of breakpoints per observer, the fine coding conditions provide reasonably sensitive measures.

In Experiment 1, this alignment effect was shown to be more substantial for familiar activities than for unfamiliar ones. We hypothesized that this

---

[2] We would like to thank an anonymous reviewer for suggesting this analysis.

difference reflects the influence of hierarchically organized event schemata, which require prior exposure to an activity to form. The results of neither Experiment 2 nor Experiment 3 support this hypothesis. However, Experiment 3 also failed to provide a statistically reliable replication of the original familiarity effect (though there were trends consistent with the effect). This makes interpretation of the relationship between expertise, familiarity, and the alignment effect difficult to interpret. The failure to observe the predicted interactions could be taken as evidence inconsistent with the hierarchical bias hypothesis. However, it could also be that the familiarity effect observed in Experiment 1 is simply not substantial enough to burden with additional experimental manipulations.

In retrospect, it seems likely that the unfamiliar event was simply not unfamiliar enough. Even though the novices in these studies did not know how to assemble a saxophone, they have had extensive experience assembling other objects. It could be that the meta-knowledge about assembling objects facilitates interpretation of assembling a saxophone and other unfamiliar objects, diminishing effects of instruction and experience in segmentation.

## Experiment 4: Structure in Memory for Events

Patterns in perceptual encoding in the previous experiments suggested the influence of hierarchically organized schemata for events on the perception of ongoing activity. If such cognitive structures guide perception, they should surely influence memory. The notion that schemata for events guide memory goes back at least to Bartlett (1932). As was described previously, there is extensive evidence that hierarchically organized event schemata guide recall for

written stories (Abbott et al., 1985; Bower et al., 1979; Rumelhart, 1977) and videotapes of live-action events (Brewer & Dupree, 1983; Lichtenstein & Brewer, 1980). It is therefore natural to expect to find a relationship between the segmentation and linguistic data obtained in these perceptual experiments and later recall for the activities presented.

In one regard, one should expect memory for events to preserve the structure observed in the perceptual experiments. The differences in Experiment 1 between descriptions of coarse, boundary-fine, and internal-fine units suggested the influence of hierarchically organized schemata. Given such an encoding bias, one would expect these differences to be preserved in memory. We therefore expected to see similar differences between coarse and fine units in memory as had been seen in perception. The primary motivation of this experiment was to compare the syntactic and semantic content of event descriptions from memory, under fine and coarse coding conditions, to the descriptions obtained during on-line segmentation. To preserve as rich a representation to draw on as possible, we chose to study memory under very short delay conditions, thus focusing on the difference between narrating a live action and reporting on one in the immediate past.

In another regard, one might expect to see differences in the level at which participants described activity in memory and perception. When recalling stories from memory, people sometimes shift to a coarser grain of description, omitting the lower levels of the partonomy (Rumelhart, 1977). However, given the short delay in the current experiment, such effects might not have time to exert an influence.

In short, this experiment was designed to test the hypothesis that differences between fine-unit and coarse-unit descriptions from memory would replicate those of descriptions while viewing activity. We also were interested in examining possible changes in level of segmentation from memory.

## Method

### Participants

22 Stanford students participated in this experiment in partial fulfillment of a course requirement.

### Materials and Procedures

The stimuli were identical to those in Experiment 1: 8 videotapes were used, making up performances of all four activities by the two actors.

The procedure was adapted from that of Experiment 1 with one important modification: Participants described event segments by writing them down immediately after watching the videotape, rather than describing them as they watched. They performed no perceptual segmentation.

Participants were run in groups of 1 to 8. Upon arriving, the experimenter explained to them that they would be watching a videotape of a person engaged in an activity. The experimenter explained that after watching the activity, they would be asked to divide the behavior of the person into the smallest units (for the fine-unit coding condition) or largest units (for the coarse-unit coding condition) that seemed natural and meaningful to them. They were told that they would be given a piece of paper and asked to write down, for each unit, what happened in that unit, and asked to use a separate line for each unit. They watched the example tape and recorded their units on paper as instructed.

They then did the same with the four experimental activities (arranged in one of four different orders, depending on group). Next, they participated in an unrelated experiment for approximately 20 minutes. Then they watched the four experimental activities again. This time, if they had been asked to segment the activity into the smallest units that felt natural and meaningful before, they were asked to now segment into the largest units that felt natural and meaningful, and vice versa.

Results

Of the 22 participants, data from 5 were not usable. 2 failed to complete the experiment. While most participants gave concrete descriptions of the activity in the videotapes, 2 gave unresponsive descriptions that were deemed unanalyzable. (Examples: "desperate attempt of a man to ascertain his own influence over the world and to combat the forces of chaos (I'm serious)," and "Atmosphere: somewhat bright light in the bathroom.") A single participant failed to follow the instruction to record units one-per-line. Of the remaining 17 participants, 9 were given the fine-unit coding instructions before the first viewing of the videotapes and the coarse-unit instructions for the second; for the other 8 participants the opposite order was employed.

Two participants, when writing coarse-unit descriptions, spontaneously organized the descriptions hierarchically, using numbering and indentation to indicate the partonomic relationships. For these lists, the fine units were deleted prior to analysis.

The written descriptions were transcribed and analyzed using the same procedure as was used in Experiment 1. Again, syntactic features were coded by two judges and disagreements were adjudicated by the first author. Polysemy

counts were obtained using WordNet (Fellbaum, 1998, version 1.6), and the norms for generality of verbs and nouns, and goal-directedness of verbs, were taken from Experiment 1.

---

Insert Figure 12 about here

---

One important difference between the design of this experiment and the on-line segmentation study is that the memory design does not allow the recording of temporal locations for unit boundaries. Because the temporal locations of participants' unit boundaries were unknown, we were unable to parcel the fine-unit descriptions into boundary and internal units, as in Experiment 1. Therefore, only coarse- and fine-unit descriptions were compared.

To a striking degree, the patterns in the linguistic features of objects replicated those of the on-line segmentation data. Again, utterances were predominantly telegraphic descriptions of simple intentional acts. On the whole, as before, reference to objects were more specific at the coarse level and references to actions were more specific at the fine level. The data are summarized in Figure 12. Object ellipsis was more likely for fine units than coarse units. The same was true for object recurrence and repetition. There were on average more objects per utterance in fine units than in coarse units. Fine-unit objects were more polysemous than coarse-unit objects. Objects used in fine-unit descriptions were rated as slightly less general than those in coarse-unit descriptions. All of these patterns replicate those from on-line descriptions. Objects were more slightly likely to be pronominalized in coarse units than in fine units, a pattern that differed from the on-line segmentation results. Object pronominalization was also more likely overall than in Experiment 1.

For verbs, the syntactic features replicated the patterns observed in the on-line segmentation task. There were no instances of verb ellipsis for the coarse units, but it did occur occasionally in the fine units. Verbs were slightly more likely to be marked as recurrent under fine-unit coding conditions, and there were essentially no differences between coarse and fine units for the use of the progressive form. The pattern for semantic features of verbs in descriptions from memory, however, was exactly opposite to that for on-line descriptions. Fine-unit verbs were more polysemous than coarse-unit verbs, rated as more general, and rated as more goal-directed.

Grammatical subjects of event descriptions from memory showed the same patterns as did subjects in on-line descriptions. As in Experiment 1, subjects were usually elided, and this was especially true under fine-unit coding conditions. The subject of the utterance was more likely to be pronominalized under coarse-unit coding conditions than fine-unit coding conditions, though the difference (as well as the overall base rate) was small.

---

Insert 14 and Figure 14 about here

---

Because event segments were written from memory rather than produced by on-line segmentation, this design provides no direct record of the length of time taken up by each unit. However, it was possible to estimate mean unit lengths in a straightforward fashion. For each viewing by a given participant of a given tape, the number of recorded units was counted. This number was then divided into the length of the tape in seconds, giving an estimated mean unit length. The results are compared to the directly measured unit lengths from Experiment 1 in Table 2. Both coarse and fine units were

somewhat longer (i.e. fewer) when produced from memory than when generated by on-line segmentation. (Outliers were removed as described previously.)

---

Insert Table 2 about here

---

Discussion

When observers described activity from recent memory, their descriptions were quite similar to those given by observers who described activity while they watched it. The same relationships between syntactic features of objects, verbs, and subjects and segmentation level were observed, and the same relationships between semantic features of objects and segmentation level were also present. The one exception is the relationship between semantic features of verbs and segmentation level. For on-line descriptions, coarse units were more polysemous, more general, and more goal-directed than internal fine units. For descriptions from memory, the opposite was true. The general pattern support the hierarchical bias hypothesis, indicating that the structured representations that guide perception also influence memory. Another striking (though anecdotal) piece of evidence comes from the fact that three participants in this study spontaneously produced hierarchical descriptions in this experiment, though explicitly instructed not to. (As noted, one participant's data were corrected before analysis. The other participants were two of the group whose data had to be excluded.) It is as if, for these participants, linear list of events void of hierarchical structure did not "count" as good descriptions of the activity.

To what might the differences between verb semantics in memory and in perception be due? One possibility is that they reflect the different production constraints of verbal and written descriptions. Another possibility is that they reflect re-coding of the coarse units, from a representation closely tied to the physical activity involved to a more schema-influenced representation that is less specific about the physical actions and more related to the goals and plans of the actor

Overall, objects and verbs produced from memory were similar to those produced on-line in their semantic content: Both verbs and objects were similar in their overall polysemy and generality, and verbs were similar in their overall degree of goal-directedness. (Compare Figure 7 to Figure 12, and Figure 8 to Figure 13.)

There was some evidence of schema-based consolidation in this experiment, reflected as a shift in the segmentation level from a finer grain to a coarser one. Observers produced fewer units from memory than during on-line segmentation. However, these differences may reflect constraints of writing, as compared to speaking, as well as effects of memory per se. The effects of medium are difficult to assess. On the one hand, writing certainly requires more effort than speaking. On the other hand, participants in the current memory experiment were under no time pressure in producing their descriptions, while participants in the segmentation experiments were constrained by the need to keep up with the activity as it happened. A better understanding of consolidation in event memory will require studies in which the output medium is matched and the delay period is parametrically varied.

Comparing both the overall characteristics of descriptions, and the differences between coarse-units and fine-unit descriptions, there is a striking similarity between descriptions from memory and from perception. This similarity suggests that the segmentation structure of the activity at encoding is playing an important role in memory. This suggestion is supported by recent work examining the relationship between cues to segmentation structure and later memory. In one experiment, Boltz (1992) showed observers a dramatic movie, with commercials placed so as to either reinforce or obscure the natural hierarchical structure. When commercials supported the natural organization, recall memory for the drama and recognition memory for scene order in the drama were improved. Placement of temporal breaks that supported the natural structure also improved memory for the total duration of the movies; a similar effect obtained for memory for the duration of musical selections (Boltz, 1995).

## Experiment 5: Comprehending Event Communications

Experiment 1 provided evidence that participants' fine-grained segmentation of ongoing activity contained embedded within it a representation of the part-whole relations of the fine units to larger units. This was evident perceptually, in the alignment effect, and linguistically, in the differences between internal and boundary fine units. The latter is essentially a correlation between the temporal structure of the activity and the linguistic features of the utterances. If a correlation exists such that it can be detected with the relatively crude coding systems employed here, might readers of descriptions of events also be sensitive to these linguistic features—and perhaps others? On the other hand, the

descriptions observers gave were not of the sort that occur in typical discourse directed at others (Clark, 1996). Their utterances were far more elliptic and contained little if any bridging or background information.

Nevertheless, human readers of descriptions of events have a vast store of background knowledge to bring to bear in inferring structure from even such telegraphic descriptions of events. A reader drawn from the same population as the participants in Experiment 1 is aware not just of syntactic features and general semantic features such as generality and goal-directedness, but also of the specific semantic structure of the domains. For example, the rating system employed here only "knows" that the word "dishwasher" is relatively specific (1.60 on a 1-5 scale in the ratings)—but the least domestic of undergraduates knows what one does with a dishwasher, even without of first-hand experience.

Based on these considerations, we predicted that participants would be able to extract to some extent the hierarchical structure exhibited in the segmentation data of Experiment 1, based solely on the fine-unit descriptions. This hypothesis was tested by presenting new participants with the fine-unit descriptions and asking them to group the fine units into larger units. We predicted they would group the fine units in much the same way as the on-line observers did in their coarse-unit coding.

It is worth noting a few features of this task and the materials used. Recall that the original participants were instructed to segment the activity into natural and meaningful units, and then describe each unit after tapping a key to mark its end—a highly non-conversational task. Further, they spoke to a tape recorder rather than to another human being. Also, they described coarse and fine units on separate viewings, so there was no opportunity to compare the segmentation

or descriptions. Finally, the readers in the current experiment were faced with the task of extracting structure for a simple, poor representation of an activity: a list of the transcribed event descriptions, and no more. All of these features make the readers' structure extraction task more difficult; nevertheless, we thought the linguistic features and background knowledge available to readers would be able to overcome these challenges.

## Method

### Participants

23 Stanford students participated in this experiment in partial fulfillment of a course requirement.

### Materials

The materials for this study were constructed from transcripts of the fine-unit event descriptions of participants in Experiment 1. All of the transcripts were printed on 11" x 17" paper. (The large-format printing was necessary to accommodate the longest transcripts on a single page with readable type. For one especially long transcript, two pieces of paper were taped together to make the stimulus page.) Each transcribed utterance appeared on the left side of the page. A heavy vertical line marked off the blank right side of the page. A heading over the right side said "Write your descriptions here."

### Procedure

Participants were informed of the source of the transcripts. They were instructed to divide the list of utterances into groups and then label each group:

> We would like to know how the individual items fit into larger groups.
>
> Look at the list and decide how to divide it into groups. All the

groups should be continuous. Mark your grouping by drawing a line between each of the groups. You can make as many or as few groups as you like. There is no right or wrong answer.

For each group, think of a description of what is happening in the whole group. Write the description of each group to the right of the set of items.

After receiving these instructions, participants were given a transcript form was selected at random from those available. As each form was completed, the experimenter selected a new form at random. Once a complete set of transcripts had been used, a new set was generated. This experiment was conducted during a prescribed period of time (to fill a delay in an unrelated memory experiment), so this procedure was continued until the 15-minute delay period was finished. Participants completed between one and nine transcripts (median = 3.00). Data from one participant was unusable because the instructions were not followed.

Results

By dividing the fine-unit transcripts into groups, participants essentially created a new set of coarse-unit breakpoints, which we will call "grouped-unit breakpoints." Each of the fine units in the transcript was scored as a grouped-unit breakpoint if it was the last unit in the marked groups. Corresponding to each fine-unit description in the transcript is the location in time at which the participant in the original experiment tapped the key. These tap times were used as an estimate of the temporal location of the grouped-unit breakpoints. (Because the forms were distributed by random selection and one participant's forms were unusable, there were zero, one or two sets of coarse breakpoints per transcript.)

The locations of the grouped-unit breakpoints give the means to compare the event structure extracted solely from the verbal transcripts with that perceived in the on-line segmentation task. The original participants in the segmentation study (Experiment 1) generated two sets of breakpoint locations for each viewing: one for coarse units and one for fine-units. Participants in the current experiment generated a new set of coarse-unit breakpoints: the grouped-unit breakpoints. To compare the two structures, we calculated for each of the grouped-unit breakpoints the distance to the nearest coarse-unit breakpoint from the corresponding viewing. The distribution of those distances is plotted in Figure 15. To the extent that readers of fine-unit descriptions were recovering the same structure as that perceived by the authors of those descriptions when they viewed the activity, two things should be the case. First, grouped-unit breakpoints should be on average close to coarse-unit breakpoints. Second, grouped-unit breakpoints should be unbiased relative to coarse-unit breakpoints; that is, the former should neither lead nor lag the latter. Both predictions held.

We evaluated the first hypothesis, that grouped-unit breakpoints should be close to coarse-unit breakpoints, using an analog of the continuous analysis from Experiment 1. The distance from each grouped-unit breakpoint to the nearest coarse-unit breakpoint from the same viewing was calculated. All the distances from a single viewing were then averaged to compute a mean score. If participants in the current experiment chose fine units to mark as grouped units in a fashion that was independent of the relationship between fine and coarse perceptual units, the expected value of this score is the mean of the distance from each of the fine-unit breakpoints in that viewing to the nearest coarse-unit

breakpoint. This was calculated for each viewing. On average, the observed mean-distance scores were reliably less (mean = 8190 ms, SEM = 2020) than the calculated expectation (mean = 12600 ms, SEM = 2480), $t(52) = 5.05$, $p < .001$.

---

Insert Figure 15 about here

---

We evaluated the second hypothesis by simply comparing the distribution of obtained distances between grouped-unit and coarse-unit breakpoints to zero, the unbiased expectation. While the mean of the distribution, at 2240 ms, was reliably higher than 0, $t(435) = 2.19$, $p = .03$, one can see that this difference is quite small both in absolute terms and relative to the spread of the distribution (its standard deviation as 21400 ms). Thus, while the grouped-unit breakpoints did lag the original coarse-unit breakpoints on average, this lag was relatively small.

Overall, participants in this study chose fewer breakpoints (i.e. larger units) than participants who performed the on-line segmentation under coarse-unit coding conditions in Experiment 1. The mean number of breakpoints per viewing in the current study was 5.52 (SEM = .390), as compared to 9.18 (SEM = .859) during on-line segmentation. Calculating breakpoint locations as described above, this led to longer breakpoints in for the grouping task (mean = 57.4 sec, SEM = 5.78) than for the on-line segmentation task (mean = 34.3 sec, SEM = 2.61). These differences may reflect differences in the effort required to write, as opposed to speak, unit descriptions; they may also reflect the lack of richness of the transcript stimuli relative to the films.

Discussion

In this experiment, participants were given a simple list of telegraphic descriptions of event segments. These lists had been generated by other participants while observing videotapes of ongoing activity. The original observers were instructed to simply segment and describe the events as they happened, not to provide any structure, hierarchical or otherwise. Furthermore, they had been placed in the highly non-conversational situation of talking into a tape recorder. Despite the impoverished nature of the descriptions, readers of these transcripts were able to extract structure from them in a manner that replicated with high fidelity that generated by the original observers. When asked to group the fine-unit descriptions in the transcripts into larger units, the boundaries participants generated were quite close to the boundaries generated by the original observers under coarse-unit coding instructions, and had only the slightest tendency to lag the original breakpoints.

These results suggest that readers in this experiment were able to extract the perceptual structure of the activity as it happened based only on these simple transcripts. On what grounds could they do this reconstruction? One obvious source of constraint is background knowledge about the activities involved. Another possible source of information is the linguistic differences between internal and boundary breakpoints uncovered in Experiment 1. It may be that readers of narrations, even highly simplified ones such as these, are sensitive to the syntactic and semantic cues that speakers use to embed information about event structure within running linear descriptions. It is possible that readers employ a version of the hierarchical bias hypothesis to decode texts: They assume that the writer's cognitive representation is hierarchically structured and

that the text will reflect this in its syntactic and semantic structure. This heuristic then guides the formation of the reader's cognitive representation of the described activity.

## General Discussion

Hierarchical structure

In our view, the principle significance of the results presented here is this: These data indicate that the same cognitive structures that have been hypothesized to guide story understanding, memory for events, planning for future activity, and understanding of one's own past actions guide perception of ongoing activity as it happens. We have described these cognitive structures as event schemata; they are closely related to plans, story grammars, and scripts. The data argue strongly for the hierarchical bias hypothesis: Observers' perception of event structure is biased by the influence of hierarchically-organized schemata for recurring events.

In Experiment 1, it was found that observers of everyday activity were disposed to organize the activity in terms of partonomic hierarchies. Participants were asked to segment everyday activities while watching them. Each participant segmented each activity twice, once under fine-unit and once under coarse-unit coding instructions. Within individuals, the boundaries of the coarse units tended to be closer to the boundaries of fine units than predicted by chance. This alignment effect was mediated by the familiarity of the activity, and was more pronounced when participants described the activity while segmenting it than when they only performed the segmentation. The alignment effect was replicated in Experiments 2 and 3. However, attempts to affect segmentation

behavior by manipulating familiarity either by instruction or by selection of experts were unsuccessful, perhaps due to the relative weakness of the familiarity effect, perhaps due to a meta-schema for putting things together.

A bias toward hierarchical structure was also observed in the descriptions observers gave of activity. In Experiment 1, descriptions of coarse and fine units differed in their syntactic structure and semantic content. Hierarchical structure was observed in the fine unit descriptions that recapitulated the perceptual alignment effect. Descriptions of fine units that were near the boundaries of that viewer's coarse units were more similar to the coarse-unit descriptions than were the remaining fine-unit descriptions. The differences between fine- and coarse-unit descriptions were by and large preserved in descriptions from memory (Experiment 4). This suggests that some of the effects of schemata on memory for events may be directly due to encoding processes, rather than post-event re-coding. In Experiment 5 it was found that the structure in an observer's descriptions of fine units, plus readers' background knowledge, was sufficient for readers to extract the relationship of those fine units to the original observer's coarse units with high fidelity. This implies not only that describers of activity embed information about the structure of the activity in running descriptions, but that readers are highly sensitive to that information.

Structure From the World and From the Mind

One might ask: To what extent do they reflect facts about the way the world is structured rather than insights into how the mind is organized? To begin with, might it be the case that the alignment effect of Experiments 1-3 reflects something simple about the nature of the activities and/or the design of the experiments, rather than the deep structure of the cognitive architecture?

Two obvious alternatives suggest themselves. (1) It could be that on their second viewing of a given activity, participants simply recalled where they tapped before and were disposed to tap there again (the "memory hypothesis"). (2) It could be that participants identified breakpoints by simply looking for peaks in the level of some continuously varying feature or features in the physical activity, and the segmentation level served just to manipulate the threshold of this peak-finding algorithm (the "threshold hypothesis"). Both the memory hypothesis and the threshold hypothesis have the attractive property of being parsimonious. They unfortunately cannot account for many of the effects observed here. If participants rely on memory for the prior viewing to reproduce segment boundaries on the second viewing, increasing attentional load should impair this memory retrieval. Thus, the memory hypothesis would have to predict that participants would show less of an alignment effect while describing the activity than while simply segmenting it—the opposite of what occurred. Moreover, the memory hypothesis cannot account for the finding of an alignment effect in the between-participants analysis of the first-viewing data. If participants are simply monitoring physical features of the activity and segmenting on the basis of those features, the familiarity of the activity should have no effect on this process. Thus, the threshold hypothesis would have to predict (again incorrectly) the absence of a familiarity effect. Finally, neither the memory hypothesis nor the threshold hypothesis offer a coherent account of the linguistic effects found in Experiments 1, 4 and 5. For the present, then, the hierarchical bias hypothesis seems to be the most compelling explanation of the results obtained in these experiments. However, this does not imply that top-down biases on perception are the whole story. The world presents organized

patterns of physical activity which can also guide event perception in a complementary, stimulus-driven fashion.

Linking perception to function

The descriptions of both coarse and fine units consisted of actions on objects, that is, of achievements or accomplishments, not of activities or states. At both coarse and fine levels, language reflected intentions and goals. Although alignment of coarse and fine units was greater under description than under simple viewing, there was a high degree of alignment even under simple viewing, implying that a partonomic structure of intentions and goals underlies perception of ongoing events.

What is the path from perception of breakpoints to imputation of intentionality? Newtson, Engquist and Bois (1977) asked participants to segment activities at either a fine grain, a coarse grain, or without specifying the segmentation level. They then coded the stimuli using a dance notation, which provided a discrete coding for the position of the bodies of the actors over time, once per second. This coding allows computation of a change score that represents how much movement is occurring at each point in time. They singled out points in time that a large number of participants had identified as breakpoints and compared these to non-breakpoints. Transitions in and out of breakpoints were characterized by unusually large change scores. This result held strongly for fine and natural units, and weakly for coarse units. So, at least for fine-grained perceptual segments, natural units correspond to locations in time at which an objective feature of the stimulus, the amount of biological motion, is at a peak. These physical features, then, correspond to a

psychologically natural division of ongoing activity into maxima of dynamic change (segment boundaries) and relative static periods (segments)

The breakpoints of dynamic change in events may signal changes in kind and function as well. Objects are spontaneously segmented into parts by changes in contour (Biederman, 1987; Hoffman & Richards, 1984), such as arms or legs on people or chairs. Parts have a dual role, one in perception and one in function, as different parts are also associated with different functions (Tversky & Hemenway, 1984); legs support both people and chairs, and tops cover both upper bodies, bottles, and carrots. Categorization in children suggests that the dual role of parts in objects allows inference from appearance to function. By analogy, different event parts, easily distinguished by relative activity, may signal different event functions.

Why is it that goal relationships tend to align with physical feature changes? One explanation can be found in Michotte's (1963) studies of perceived physical causality. Michotte showed that in paradigmatic cases of perceived causality, a single motion is projected from one object onto another. This transformation of the motion is exactly a point of large changes in physical features of the situation—precisely the points in time at which observers are disposed to segment natural activity (Newtson et al., 1977). At the moment one object is influencing another, many physical features of the situation are changing. Low-level goals are often satisfied or blocked by physical interactions between objects. Another source of empirical evidence for the convergence of structural and functional information comes from a study of memory of television stories. Van den Broek and his colleagues found that the position in the hierarchical goal structure of a story predicted rates of recall (van den Broek

et al., 1996). This tendency increased from childhood to adulthood. Moreover, they found that hierarchical position of an event unit was correlated with the number of causal connections to and from it, and with its likelihood to be embedded in a causal chain. Based on regression analyses, the authors argue that causal connectedness for the most part drives the other effects.

Thus, moments at which goals are satisfied or blocked tend to be moments at which objects are interacting causally, and those moments are the ones during which the most physical features are changing. Bottom-up, perceptually-driven information about the physical features of the activity correlates with top-down, conceptually driven information about goals, plans and causation. An organism can become sensitive to these correlations through both evolution and learning. Schemata for events are precisely a distillation of these patterns of redundancy that allow the observer to fill in missing information and make inferences on a given viewing of a particular activity (Zacks & Tversky, submitted).

Qualitative Differences Between Coarse and Fine Units

Descriptions of coarse and fine units differed qualitatively. For the activities examined here, coarse-unit descriptions tended to precisely specify a set of objects with which the actor interacted, but leave vague what the particular interactions were. On the other hand, fine-unit descriptions specified the objects vaguely but were relatively precise in specifying the actions performed on them. It remains to be seen how well this pattern generalizes to other kinds of activity. It may reflect a general principle that coarse-grained actions are distinguished by the objects (or major parts of objects, as in assembling a saxophone) they involve because these correlate well with the goals of the actor(s). This principle would

account for the finding for object-action relationships emerge early, along with a focus on actors' intentions (Meltzoff, 1995; Woodward, 1998). Put another way, coarse units are punctuated by objects (and by implication, actions) and fine units are punctuated by refined actions on the same object.

These studies have focused deliberately on everyday, goal-directed activities that include one actor and a set of objects. Within that domain the stimuli used here sampled both familiar and unfamiliar activities, and these effects seem to hold for both. What about activities where there is no tightly specified goal? Attending a fair or playing at the beach come to mind. However, even entertainment activities such as these appear to contain "pockets" of goal-directed activity (winning the ring-toss, retrieving the frisbee from the water). It may be harder than it seems to find truly goal-free activity involving animate agents. If such activities can be identified, it is interesting to ask whether the mechanisms described here still apply. A goal-based event schema may simply fail on such activity, or the perceptual mechanism may be constructed such that it imputes goals where none objectively exist.

What about activities where there are multiple actors with different goals, or one actor with multiple goals? Intuitively, if the goal structure corresponding to one actor can be described as a strict hierarchy (i.e. a tree) then the goal structure for a dyad or group of actors will presumably correspond to a more complex family of directed graph. One possibility is that different event units are identified depending on the actor who forms the focus of intentions, and for any given focus a strict tree is formed. The different hierarchies would then share nodes up to some level of description, generating a MultiTree (Furnas & Zacks, 1994). Or it may be that observers track the goals of multiple actors in parallel,

generating more complex structures to describe the relations among event units. Different classes of structure may be diagnostic of cooperation, competition, and independence. Explorations of the phase relationships between event segments in multiple-actor activities are suggestive in this regard (Newtson et al., 1987). Similar issues arise when one person performs multiple activities simultaneously or one activity that satisfies multiple goals. In these cases, activity involving only one actor will require representational systems more complex than simple hierarchies. Thus, activities with multiple actors or multiple goals may be more complex than those studied here but do not seem to differ in principle.

An important related question is whether people apply the mechanisms observed here to activities in which there are no animate agents. It may be that when humans observe events like waves washing on a beach, a volcano erupting, or a rock rolling down a hillside, they apply the same event perception tools that are applied to animate agents, resulting in a perceptual "intentional stance" (Dennett, 1987).

An Object/Action Account of Event Structure Perception

With this analysis in mind, we return to the object/action account of event structure perception, first presented in discussing Experiment 1. Common events are segmented into a partonomic hierarchy, punctuated by objects or major object parts (and concomitant actions) at a higher level and by refined actions on the same object at a lower level. Descriptions of segments, both simultaneous and retrospective, indicate that different objects are associated with different higher level functions or goals whereas different actions on the same object are associated with more refined functions or goals. It is intriguing that object/action units may serve a pivotal role in event segmentation. In his

insightful exegesis of the art of comics, McCloud (1993) recorded the frequencies of different types of transitions from one frame to another in a sample of 22 well-known American, European, and Japanese comic artists. For each, the most frequent change from one frame to the next was action-to-action. Second in frequency was a change in subject, followed by a change of scene. McCloud did not examine changes in object. Notably, moment-to-moment, within-action changes were extremely rare.

Research on animals, babies, and children indicates a privileged status for interactions on objects in developing an understanding of events. Byrne (in press-a; in press-b) has proposed that underlying execution, and especially imitation, of behavior is comprehension of the hierarchical organization of behavior. However, he argues that comprehension of events rests on detecting recurring statistical patterns of units of behavior, rather than on their content. For example, elements within a module are more tightly bound together than elements between modules. They may appear as a unit in different activities and they are less likely to be interrupted. Byrne argues that because of these statistical regularities in behavior, hierarchical structure of events may be extracted without imputation of causality or intentionality. This analysis shares reliance on statistical properties of the input with the proposals of Avrahami and Kareev (1994) about events, with those of Saffran and her colleagues (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Saffran, Newport, Aslin, Tunick, & et al., 1997) on language, and with those about the role of correlated features in object categories by Rosch (1978).

However, the actual case studies on animals, from black rats to gorillas, suggest that event segmentation has more to go on than just the statistical

properties of event units. Qualitative as well as quantitative information is available and seems to be influential. Specifically, the case studies implicate objects or major parts of objects, whether nests or food or tools or other animals, as critical determinants of the junctures between segments and of the hierarchical structure as well. This view has been corroborated by laboratory studies of interactions with artificial fruits by chimpanzees and preschool children (Whiten, in press; Whiten & Custance, 1996). Careful comparisons have shown that children are especially sensitive to the hierarchical structure of events independent of sequential structure. Importantly, the correspondence of event segments with actions on objects seems to allow inferences, perhaps rudimentary, of intentionality. Thus, the structure within events complements inter-event statistical relationships as a basis for event comprehension

Infants, too, appear to use objects to delineate events. Woodward (1998) trained 9 month old babies to look at a simple event consisting of an action directed toward an object. In later tests, infants looked longer when the object was switched than when the action was changed. Further research indicates that by one year of age, infants are predisposed to interpret actions on objects as goal-directed (Woodward & Sommerville, submitted). That comprehension of events is affected more by presumed goals than by details of actions is supported by research on imitation in neonates and children. Neonates modulate their own behavior, bringing it closer to the adult model that instigated it (Meltzoff & Moore, 1995). Children as young as 18 months successfully achieve a goal even after watching varying actions on objects that failed to achieve it (Meltzoff, 1995).

These findings and others have led Baldwin and Baird (1999) to argue that action analysis is central to inferring intentions, and that the links between action and intention are especially strong at natural breakpoints. Congruent with the position we put forth here, Baldwin and Baird maintain that objects are integral to comprehending actions and vice versa. Our work suggests that neither object nor action alone is sufficient for understanding events. Whether a sheet is folded or spread, whether an apple is eaten or thrown away, whether a book is opened or packed is critical to interpretation of the event. Similarly, whether a letter or a sheet is folded, whether an apple or a pill is ingested, whether a book or a drawer is opened changes the meaning of the activity. It is the interaction, the conjunction of object with action, their correlated use in behavior, all readily apparent in perception, that enables interpretations of events as goal-directed, purposeful, intentional.

Multiple Constraints on Event Understanding

In sum, the mind has access to a number sources of information about structure in activity. Further, different kinds of information are correlated. Goal-directed activity reflects the goals of actors and the constrained relationships of recurring activities. It tends to be hierarchical because goals tend to be satisfied by the recursive satisfaction of sub-goals. The goal structure of activity aligns with its physical structure because the satisfaction of goals tends to give rise to distinctive physical characteristics, particularly in the relationship of actors to objects. The distinctive physical features of causal interactions may mediate this relationship. Language in general tends to capture the features whose changes mark boundaries in activity. Information from each of these domains imposes constraints on the others.

People simultaneously keep track of physical changes, goals and plans, causes and effects, actions and objects. It is tempting to try to explain our understanding of events in terms of one of these set of features. However, the fact that each of them tells something about the others has two consequences. First, it makes tractable the problem of following what is happening a complex dynamic world. Second, it means that an account of event understanding must include multiple sources of information—and their connections.

## Author Note

Jeff Zacks is now at Washington University in St. Louis. Please address correspondence concerning this article to: Jeff Zacks, Washington University, Department of Psychology, Campus Box 1125, St. Louis, MO 63130-4899 (email: jzacks@artsci.wustl.edu). Gowri Iyer is now at the University of California at San Diego and San Diego State University. We would like to thank Gordon Bower, Herb Clark, and John Gabrieli for stimulating discussions of this work, and thank Yaakov Kareev and an anonymous reviewer for helpful comments on a previous draft. We are grateful to Perrine Bhakshay, Caroline Carter, Crosby Grant, Mike Jahr, and Dan Maidenberg for assistance in various aspects of the research. This work has benefited from the support of a National Science Fellowship and a Stanford Humanities & Sciences Dissertation Fellowship to the first author, and from some support from Interval Research Corporation. Our thanks to The Starving Musician for providing saxophones for use in these experiments.

# References

Abbott, V., Black, J. H., & Smith, E. E. (1985). The representation of scripts in memory. Journal of Memory and Language, 24, 179-199.

Abelson, R. P. (1981). Psychological status of the script concept. American Psychologist, 36, 715-729.

Anderson, S. J., & Conway, M. A. (1993). Investigating the structure of autobiographical memories. Journal of Experimental Psychology: Learning, Memory, & Cognition, 19, 1178-1196.

Avrahami, J., & Kareev, Y. (1994). The emergence of events. Cognition, 53, 239-261.

Baldwin, D. A., & Baird, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), Early social cognition (pp. 215-240). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.), Remembering reconsidered ecological and traditional approaches to the study of memory . Cambridge: Cambridge University Press.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. New York: The Macmillan Company.

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, 115-117.

Boltz, M. (1992). Temporal accent structure and the remembering of filmed narratives. Journal of Experimental Psychology: Human Perception & Performance, 18, 90-105.

Boltz, M. G. (1995). Effects of Event Structure On Retrospective Duration Judgments. Perception & Psychophysics, 57, 1080-1096.

Bower, G. (1982). Plans and goals in understanding episodes. In A. Flammer & W. Kintsch (Eds.), Discourse Processing (pp. 2-15). Amsterdam: North-Holland Publishing Company.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Brewer, W. F., & Dupree, D. A. (1983). Use of plan schemata in the recall and recognition of goal-directed actions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 117-129.

Byrne, R. W. (in press-a). Imitation without intentionality: Using string parsing to copy the organization of behaviour. Animal Cognition.

Byrne, R. W. (in press-b). Seeing actions as hierarchically organized structures: Great ape manual skills. In A. Meltzoff & W. Prinz (Eds.), The imitative mind: Evolution, development, and brain bases .

Casati, R., & Varzi, A. C. (1996). Events. Aldershot, England ; Brookfield, Vt.: Dartmouth.

Clark, H. H. (1996). Using language. Cambridge England: Cambridge University Press.

Cohen, C. E., & Ebbesen, E. B. (1979). Observational goals and schema activation: a theoretical framework for behavior perception. Journal of Experimental Social Psychology, 15, 305-329.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. Behavior Research Methods, Instruments & Computers, 25, 257-271.

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

Ebbesen, E. B. (1980). Cognitive processes in understanding ongoing behavior. In e. a. Reid Hastie (Ed.), Person memory: the cognitive basis of social perception (pp. 179-225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fellbaum, C. (1998). WordNet : an electronic lexical database. Cambridge, Mass: MIT Press.

Foss, C. L., & Bower, G. H. (1986). Understanding actions in relation to goals. In N. E. Sharkey (Ed.), Advances in cognitive science (Vol. I, pp. 94-124). Chichester: Ellis Horwood, Ltd.

Franklin, N., & Bower, G. H. (1988). Retrieving actions from goal hierarchies. Bulletin of the Psychonomic Society, 26, 15-18.

Furnas, G. W., & Zacks, J. (1994). Multitrees: enriching and reusing hierarchical structure. In B. Adelson, S. Dumais, & J. Olson (Eds.), Human factors in computing systems: CHI '94 conference proceedings (pp. 330-336). Boston: ACM.

Galambos, J. A. (1982). Normative studies of six characteristics of our knowledge of common activities (14). New Haven: Yale University Cognitive Science Group.

Goldberg, A. (1995). Constructions: A construction grammar approach to argument structure. Chicago: University of Chicago Press.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. Cognition, 18, 65-96.

Hudson, J. A. (1988). Children's memory for atypical actions in script-based stories: Evidence for a disruption effect. Journal of Experimental Child Psychology, 46, 159-173.

Kieras, D. E. (1988). What mental model should be taught: Choosing instructional content for complex engineered systems. In J. Psotka, L. D. Massey, & S. A. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 85-111). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Lichtenstein, E. D., & Brewer, W. F. (1980). Memory for goal-directed events. Cognitive Psychology, 12, 412-445.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 9, 111-151.

Massad, C. M., Michael, H., & Newtson, D. (1979). Selective perception of events. Journal of Experimental Social Psychology, 15, 513-532.

McCloud, S. (1993). Understanding comics: The invisible art. Northampton, MA: Kitchen Sink Press.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology, 31, 838-850.

Meltzoff, A. N., & Moore, M. K. (1995). Infants' understanding of people and things: From body imitation to folk psychology. In J. L. Bermudez, A. J.

Marcel, & N. Eilan (Eds.), The body and the self (pp. 43-69). Cambridge: MIT Press.

Michotte, A. E. (1963). The perception of causality. New York: Basic Books.

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. Computational Linguistics, 14, 15-28.

Narayanan, S. (1997). Talking the talk *is* like walking the walk: A computational model of verbal aspect, 19th Annual Meeting of the Cognitive Science Society (pp. 548-553). Stanford, California: Ablex.

Nelson, K., & Gruendel, J. (1986). Children's scripts. In K. Nelson (Ed.), Event knowledge: Structure and function in development (pp. 21-46). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. Journal of Personality and Social Psychology, 28, 28-38.

Newtson, D. (1993). The dynamics of action and interaction. In L. B. Smith & E. Thelen (Eds.), A dynamic systems approach to development: applications (pp. 241-264). Cambridge, MA: MIT Press.

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

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

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. Special Issue: Cognition and action. Social Cognition, 5, 191-237.

Pustejovsky, J. (1991). The syntax of event structure. Special Issue: Lexical and conceptual semantics. Cognition, 41, 47-81.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (Eds.), Representation and understanding; studies in cognitive science. (pp. 211-236). New York: Academic Press.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. Laberge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension (pp. 265-303). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. Journal of Memory & Language, 35, 606-621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & et al. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. Psychological Science, 8, 101-105.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals, and understanding: an inquiry into human knowledge structures. Hillsdale, N.J.: L. Erlbaum Associates.

Talmy, L. (1975). Semantics and syntax of motion. In J. P. Kimball (Ed.), Syntax and Semantics (Vol. 4, pp. 181-238). New York: Academic Press, Inc.

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9, 77-110.

Trabasso, T., & Stein, N. L. (1994). Using goal-plan knowledge to merge the past with the present and the future in narrating events on line. In M. Haith (Ed.), The Development of future-oriented processes (pp. 323-349). Chicago: University of Chicago Press.

Travis, L. L. (1997). Goal-based organization of event memory in toddlers. In P. W. v. d. Broek, P. J. Bauer, & T. Bovig (Eds.), Developmental spans in event comprehension and representation: Bridging fictional and actual events (pp. 111-138). Mahwah, NJ: Lawrence Erlbaum Associates.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. Journal of Experimental Psychology: General, 113, 169-193.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. Psychological Review, 94, 3-15.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. Child Development, 67, 3010-3028.

Whiten, A. (in press). The imitator's representation of the imitated: ape and child. In A. Meltzoff & W. Prinz (Eds.), The imitative mind: Evolution, development, and brain bases .

Whiten, A., & Custance, D. (1996). Studies of imitation in chimpanzees and children. In C. M. Heyes & B. G. Galef (Eds.), Social learning in animals: The roots of culture. (pp. 291-318). San Diego: Academic Press Inc.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69, 1-34.

Woodward, A. L., & Sommerville, J. A. (submitted). Twelve-month infants interpret action in context. .

Zacks, J., & Tversky, B. (submitted). Event structure in perception and cognition. .

# Appendix A: Norms for Everyday Activities

As a precursor to preparing stimuli for the event segmentation experiments, we collected norms for a collection of 45 activities. Activities were selected to be goal-directed, to involve one actor, and involve interactions with objects. Activities were also selected that could reasonably make up part of daily life for our target research subjects (college undergraduates in the U.S), that were short, and that could be reasonably videotaped . Each was rated for familiarity, for knowledge of the steps involved, and for frequency of performance.

Method

Beginning with a list of activities from a set of norms collected by Galambos (1982), we generated a list of 45 activities that satisfied several criteria. The activities were all goal-directed, performed by a single person, and involved interactions with objects. We attempted to select activities that could reasonably occur on a given day for our target participant population, were relatively short, and could be reasonably videotaped.

The sampled activities were assembled into questionnaires that asked one of 3 questions about each of the 45 activities. The questions were (following Galambos):

- How familiar are you with each of the following activities? [How Familiar]
- How frequently do you do each of the following activities? [How Frequent]
- How well do you know the steps in each of the following activities? [Know Steps]

Each questionnaire contained one of these questions, followed by a 1-9 scale and instructions how to use it, followed by the 45 activities, listed in random order. For each question, 2 different random orders were generated, giving 6 different questionnaires.

Booklets were printed on 8.5" x 11" paper. Equal numbers of the 6 forms were assembled into booklets with other questionnaires and distributed to students in an introductory psychology class at Stanford University. Students participated in the study in partial fulfillment of a course requirement.

Results

Of the "How Familiar" questionnaires, 33 were returned; of the "How Frequent" questionnaires, 37 were returned; and 35 of the "Know Steps" questionnaires were returned. Respondents occasionally failed to answer one or more of the questions; for each question, at most 2 participants failed to respond for any of the activities queried.

---

Insert Table 1 about here

---

Ratings for the 3 questions were highly correlated, as can be seen in Table 1. The full results of the norms follow.

1. Ratings of Familiarity

| | Mean | Median | SD | SEM | Count | Min. | Max. |
|---|---|---|---|---|---|---|---|
| assembling a lamp | 4.515 | 4 | 2.224 | 0.387 | 33 | 1 | 9 |
| assembling a saxophone | 3.03 | 2 | 2.069 | 0.36 | 33 | 1 | 9 |
| brewing some tea | 6.848 | 7 | 2.21 | 0.385 | 33 | 2 | 9 |
| brushing teeth | 8.818 | 9 | 0.769 | 0.134 | 33 | 5 | 9 |
| changing a flat | 4.727 | 5 | 2.169 | 0.378 | 33 | 1 | 9 |
| changing a lightbulb | 8.303 | 9 | 1.334 | 0.232 | 33 | 4 | 9 |
| checking phone messages | 8.485 | 9 | 1.302 | 0.227 | 33 | 3 | 9 |
| doing the dishes | 8.424 | 9 | 1.501 | 0.261 | 33 | 2 | 9 |
| doing the ironing | 7.455 | 8 | 2.032 | 0.354 | 33 | 2 | 9 |
| eating some cereal | 8.697 | 9 | 0.984 | 0.171 | 33 | 4 | 9 |
| fertilizing houseplants | 4.848 | 5 | 2.412 | 0.42 | 33 | 1 | 9 |
| filing some papers | 7.576 | 9 | 2.031 | 0.354 | 33 | 2 | 9 |
| folding the laundry | 8.515 | 9 | 1.121 | 0.195 | 33 | 4 | 9 |
| grinding coffee | 5.788 | 7 | 2.342 | 0.408 | 33 | 2 | 9 |
| hanging a picture | 7.394 | 9 | 2.03 | 0.353 | 33 | 3 | 9 |
| installing a computer | 4.545 | 3 | 2.575 | 0.448 | 33 | 1 | 9 |
| juicing oranges | 6.485 | 7 | 2.152 | 0.375 | 33 | 3 | 9 |
| leaving a phone message | 8.424 | 9 | 1.324 | 0.23 | 33 | 4 | 9 |
| making popcorn | 8.061 | 9 | 1.478 | 0.257 | 33 | 4 | 9 |
| making the bed | 8.455 | 9 | 1.394 | 0.243 | 33 | 3 | 9 |
| making a campfire | 5.545 | 5 | 2.489 | 0.433 | 33 | 2 | 9 |
| making a sandwich | 8.844 | 9 | 0.515 | 0.091 | 32 | 7 | 9 |
| making coffee | 6.758 | 7 | 2.437 | 0.424 | 33 | 2 | 9 |
| paying a bill | 7.758 | 9 | 2.208 | 0.384 | 33 | 2 | 9 |
| pitching a tent | 5.758 | 5 | 2.634 | 0.459 | 33 | 2 | 9 |
| planting some seeds | 6.091 | 7 | 2.429 | 0.423 | 33 | 1 | 9 |
| playing a video game | 7.485 | 9 | 2.017 | 0.351 | 33 | 3 | 9 |
| playing some solitaire | 7.485 | 9 | 2.063 | 0.359 | 33 | 3 | 9 |
| playing some tennis | 6.121 | 6 | 2.434 | 0.424 | 33 | 1 | 9 |
| reupholstering a chair | 3.091 | 3 | 1.926 | 0.335 | 33 | 1 | 9 |
| setting up a volleyball net | 5.848 | 6 | 2.123 | 0.37 | 33 | 2 | 9 |
| sewing a button | 6.424 | 7 | 2.047 | 0.356 | 33 | 2 | 9 |
| showing slides | 5.848 | 6 | 2.252 | 0.392 | 33 | 2 | 9 |
| smoking a pipe | 4 | 3 | 2.194 | 0.382 | 33 | 1 | 9 |
| taking a photograph | 8 | 9 | 1.436 | 0.25 | 33 | 4 | 9 |
| taking a run | 7.909 | 9 | 1.91 | 0.332 | 33 | 2 | 9 |
| taking out the garbage | 7.818 | 9 | 2.038 | 0.355 | 33 | 2 | 9 |
| tying shoes | 8.848 | 9 | 0.619 | 0.108 | 33 | 6 | 9 |
| using a vending machine | 8.697 | 9 | 0.81 | 0.141 | 33 | 6 | 9 |
| using an ATM | 8.061 | 9 | 1.657 | 0.288 | 33 | 3 | 9 |
| vacuuming the floor | 8.515 | 9 | 1.326 | 0.231 | 33 | 3 | 9 |
| walking the dog | 7.091 | 7 | 1.958 | 0.341 | 33 | 3 | 9 |
| wrapping a gift | 8.303 | 9 | 1.531 | 0.266 | 33 | 4 | 9 |
| writing a letter | 8.515 | 9 | 1.064 | 0.185 | 33 | 5 | 9 |
| xeroxing a page | 8.545 | 9 | 1.034 | 0.18 | 33 | 5 | 9 |

## 2. Ratings of Frequency of Performance

| | Mean | Median | SD | SEM | Count | Min. | Max. |
|---|---|---|---|---|---|---|---|
| assembling a lamp | 2.056 | 2 | 1.372 | 0.229 | 36 | 1 | 8 |
| assembling a saxophone | 1.389 | 1 | 0.964 | 0.161 | 36 | 1 | 4 |
| brewing some tea | 4.5 | 5 | 2.287 | 0.381 | 36 | 1 | 9 |
| brushing teeth | 8.419 | 9 | 1.402 | 0.23 | 37 | 1 | 9 |
| changing a flat | 1.649 | 1 | 1.399 | 0.23 | 37 | 1 | 9 |
| changing a lightbulb | 3.667 | 4 | 1.095 | 0.183 | 36 | 1 | 7 |
| checking phone messages | 7.778 | 8.5 | 1.899 | 0.317 | 36 | 2 | 9 |
| doing the dishes | 5.694 | 6 | 1.618 | 0.27 | 36 | 2 | 8 |
| doing the ironing | 3.694 | 4 | 1.582 | 0.264 | 36 | 1 | 7 |
| eating some cereal | 6.486 | 7 | 1.693 | 0.278 | 37 | 1 | 9 |
| fertilizing houseplants | 2.027 | 1 | 1.481 | 0.243 | 37 | 1 | 6 |
| filing some papers | 5.229 | 5 | 1.699 | 0.287 | 35 | 1 | 8 |
| folding the laundry | 5.361 | 6 | 1.15 | 0.192 | 36 | 1 | 7 |
| grinding coffee | 2 | 1 | 1.867 | 0.311 | 36 | 1 | 8 |
| hanging a picture | 3.722 | 4 | 1.386 | 0.231 | 36 | 1 | 9 |
| installing a computer | 2.056 | 2 | 1.145 | 0.191 | 36 | 1 | 6 |
| juicing oranges | 2.4 | 2 | 1.649 | 0.279 | 35 | 1 | 7 |
| leaving a phone message | 6.865 | 7 | 1.669 | 0.274 | 37 | 1 | 9 |
| making popcorn | 4.083 | 4 | 1.442 | 0.24 | 36 | 1 | 7 |
| making the bed | 6.5 | 7 | 1.978 | 0.33 | 36 | 1 | 9 |
| making a campfire | 2.333 | 2 | 1.454 | 0.242 | 36 | 1 | 8 |
| making a sandwich | 5.861 | 6 | 1.334 | 0.222 | 36 | 3 | 8 |
| making coffee | 3.889 | 4 | 2.681 | 0.447 | 36 | 1 | 9 |
| paying a bill | 5 | 5 | 0.956 | 0.159 | 36 | 1 | 7 |
| pitching a tent | 2.028 | 2 | 1 | 0.167 | 36 | 1 | 4 |
| planting some seeds | 2.083 | 2 | 1.079 | 0.18 | 36 | 1 | 5 |
| playing a video game | 3.278 | 4 | 1.734 | 0.289 | 36 | 1 | 7 |
| playing some solitaire | 3.75 | 4 | 1.645 | 0.274 | 36 | 1 | 7 |
| playing some tennis | 3.314 | 3 | 1.778 | 0.301 | 35 | 1 | 7 |
| reupholstering a chair | 1.472 | 1 | 1.298 | 0.216 | 36 | 1 | 7 |
| setting up a volleyball net | 1.806 | 1 | 1.167 | 0.194 | 36 | 1 | 6 |
| sewing a button | 3.056 | 3 | 1.492 | 0.249 | 36 | 1 | 7 |
| showing slides | 1.917 | 1 | 1.402 | 0.234 | 36 | 1 | 6 |
| smoking a pipe | 1.556 | 1 | 1.107 | 0.184 | 36 | 1 | 5 |
| taking a photograph | 5.083 | 5 | 1.079 | 0.18 | 36 | 4 | 8 |
| taking a run | 5.569 | 6 | 2.129 | 0.355 | 36 | 1 | 9 |
| taking out the garbage | 5.5 | 6 | 1.404 | 0.234 | 36 | 1 | 9 |
| tying shoes | 7.667 | 8 | 2.042 | 0.34 | 36 | 1 | 9 |
| using a vending machine | 4.843 | 5 | 1.408 | 0.238 | 35 | 2 | 8 |
| using an ATM | 5.5 | 6 | 1.483 | 0.247 | 36 | 1 | 9 |
| vacuuming the floor | 4.944 | 5 | 1.286 | 0.214 | 36 | 1 | 7 |
| walking the dog | 2.429 | 1 | 2.076 | 0.351 | 35 | 1 | 9 |
| wrapping a gift | 4.111 | 4 | 0.979 | 0.163 | 36 | 1 | 6 |
| writing a letter | 5.194 | 5 | 1.737 | 0.29 | 36 | 1 | 9 |
| xeroxing a page | 5.333 | 5 | 1.454 | 0.242 | 36 | 1 | 8 |

3. Ratings of Knowledge of Steps

| | Mean | Median | SD | SEM | Count | Min. | Max. |
|---|---|---|---|---|---|---|---|
| assembling a lamp | 5.057 | 5 | 2.508 | 0.424 | 35 | 1 | 9 |
| assembling a saxophone | 3.514 | 2 | 3.091 | 0.522 | 35 | 1 | 9 |
| brewing some tea | 6.771 | 7 | 2.34 | 0.396 | 35 | 1 | 9 |
| brushing teeth | 8.543 | 9 | 0.98 | 0.166 | 35 | 5 | 9 |
| changing a flat | 3.514 | 3 | 2.716 | 0.459 | 35 | 1 | 9 |
| changing a lightbulb | 8.029 | 9 | 1.74 | 0.294 | 35 | 2 | 9 |
| checking phone messages | 8.2 | 9 | 1.587 | 0.268 | 35 | 3 | 9 |
| doing the dishes | 8.2 | 9 | 1.324 | 0.224 | 35 | 4 | 9 |
| doing the ironing | 6.371 | 7 | 2.25 | 0.38 | 35 | 1 | 9 |
| eating some cereal | 8.686 | 9 | 0.867 | 0.147 | 35 | 5 | 9 |
| fertilizing houseplants | 4.171 | 3 | 2.455 | 0.415 | 35 | 1 | 9 |
| filing some papers | 7.457 | 8 | 1.651 | 0.279 | 35 | 3 | 9 |
| folding the laundry | 8.171 | 9 | 1.071 | 0.181 | 35 | 5 | 9 |
| grinding coffee | 4.114 | 3 | 2.774 | 0.469 | 35 | 1 | 9 |
| hanging a picture | 7.114 | 7 | 2.026 | 0.342 | 35 | 3 | 9 |
| installing a computer | 3.671 | 3 | 2.342 | 0.396 | 35 | 1 | 8 |
| juicing oranges | 6.829 | 7 | 2.107 | 0.356 | 35 | 2 | 9 |
| leaving a phone message | 7.886 | 9 | 1.53 | 0.259 | 35 | 3 | 9 |
| making popcorn | 8.2 | 9 | 1.511 | 0.255 | 35 | 3 | 9 |
| making the bed | 7.943 | 9 | 1.474 | 0.249 | 35 | 3 | 9 |
| making a campfire | 5.457 | 5 | 2.683 | 0.453 | 35 | 1 | 9 |
| making a sandwich | 8.229 | 9 | 1.536 | 0.26 | 35 | 3 | 9 |
| making coffee | 6.086 | 7 | 2.628 | 0.444 | 35 | 1 | 9 |
| paying a bill | 7.857 | 9 | 1.332 | 0.225 | 35 | 5 | 9 |
| pitching a tent | 5.486 | 5 | 2.79 | 0.472 | 35 | 1 | 9 |
| planting some seeds | 5.714 | 6 | 2.346 | 0.397 | 35 | 1 | 9 |
| playing a video game | 6.143 | 7 | 2.39 | 0.404 | 35 | 2 | 9 |
| playing some solitaire | 6.914 | 8 | 2.748 | 0.464 | 35 | 1 | 9 |
| playing some tennis | 5.057 | 5 | 2.849 | 0.481 | 35 | 1 | 9 |
| reupholstering a chair | 1.914 | 1 | 1.422 | 0.24 | 35 | 1 | 5 |
| setting up a volleyball net | 5.629 | 5 | 2.302 | 0.389 | 35 | 1 | 9 |
| sewing a button | 6.286 | 7 | 2.08 | 0.352 | 35 | 2 | 9 |
| showing slides | 4.257 | 5 | 2.254 | 0.381 | 35 | 1 | 9 |
| smoking a pipe | 3.4 | 2 | 2.851 | 0.482 | 35 | 1 | 9 |
| taking a photograph | 7.486 | 8 | 1.669 | 0.282 | 35 | 4 | 9 |
| taking a run | 7.657 | 8 | 1.697 | 0.287 | 35 | 3 | 9 |
| taking out the garbage | 7.829 | 9 | 1.723 | 0.291 | 35 | 3 | 9 |
| tying shoes | 8.743 | 9 | 0.78 | 0.132 | 35 | 5 | 9 |
| using a vending machine | 8.114 | 9 | 1.43 | 0.242 | 35 | 3 | 9 |
| using an ATM | 7.857 | 8 | 1.309 | 0.221 | 35 | 4 | 9 |
| vacuuming the floor | 8.057 | 9 | 1.552 | 0.262 | 35 | 3 | 9 |
| walking the dog | 7.086 | 9 | 2.582 | 0.437 | 35 | 1 | 9 |
| wrapping a gift | 7.571 | 8 | 1.577 | 0.267 | 35 | 5 | 9 |
| writing a letter | 8.343 | 9 | 1.211 | 0.205 | 35 | 5 | 9 |
| xeroxing a page | 8.057 | 9 | 1.589 | 0.269 | 35 | 3 | 9 |

# Appendix B: Scripts for Four Selected Activities

## Familiar

MAKING A BED
take off comforter
strip the bed
get the linens
spread bottom sheet
spread top sheet
tuck in bottom
tuck in sides
put on comforter
fold back top sheet
tuck in comforter
put pillowcases on
put pillows on

DOING THE DISHES
put on apron
clean off scraps
rinse the dishes
load the glasses
load the plates
load the silverware
get the detergent
pour the detergent
put away detergent
start the dishwasher
wash off hands
put away apron

## Unfamiliar

FERTILIZING A HOUSEPLANT
get the fertilizer
get the watering-can
get the measuring spoon
measure the fertilizer
fill the can
get the plant
water the plant
return the plant
pour out excess
rinse the can
put away can
put away fertilizer

ASSEMBLING A SAXOPHONE
open the case
unpack the body
unpack the neck
remove the swab
wipe the body
attach the neck
put on the neck strap
clip on the saxophone
attach the mouthpiece
wet the reed
attach the reed
close the case

# Appendix C: Ratings of Semantic Features of Verbs and Nouns

In order to investigate the semantic content of event descriptions we collected ratings of two semantic features of verbs and one feature of nouns. For verbs, ratings of generality and goal-directedness were obtained. For nouns, ratings of generality were obtained.

Method

Words were drawn from the verbal transcripts of Experiment 1. First, objects and verbs were recorded in root form; adjectives, adverbs, and hyphenated modifiers were stripped off. Each root form was included if and only if it appeared in the verbal transcripts of 2 or more of the participants. This left a list of 123 nouns (objects) and 113 verbs. For each list two random orders were generated, and each split into 3 equal-sized sub-lists. These lists were printed on 8.5" x 11" paper, with a 5-point Likert scale next to each word and instructions at the top of the page. Each scale was labeled with the extrema of the continuum being measured.

Nouns were rated on a continuum from specific to general. The instructions for the noun rating forms were:

> In this study, we're trying to understand how nouns can differ. One way nouns can differ is in how specific or general they are. A noun like "scarf" is very specific. It describes a specific kind of noun in a way that is easy to visualize. On the other hand "clothing" is a very general noun. There are many different kinds of clothing, and it is difficult to visualize clothing in general.

Please rate each of the nouns below using the scale provided. 1 is for very specific nouns (like "scarf"). 5 is for very general nouns (like "clothing").

Two different continua were rated for verbs. On one form, verbs were (as with nouns) rated on a scale running from specific to general. The instructions for this form were:

In this study, we're trying to understand how verbs can differ. One way verbs can differ is in how specific or general they are. A verb like "slurp" is very specific. It describes precisely how the person is behaving at that moment, in a way that is easy to visualize. On the other hand "eat" is a very general verb. There are many different ways to eat, and it is difficult to visualize eating in general.

Please rate each of the verbs below using the scale provided. 1 is for very specific verbs (like "slurp"). 5 is for very general verbs (like "eat").

On the other form, verbs were rated for their goal-directedness, on a scale running from not goal-directed to goal-directed. The instructions were:

In this study, we're trying to understand how verbs can differ. One way verbs can differ is in how goal-directed they are. A verb like "complete" is very goal-directed. It strongly implies that a goal has been achieved. On the other hand, "rotate" is not very goal-directed. It could describe an ongoing process or physical event that isn't related to any goal.

Please rate each of the verbs below using the scale provided. 1 is for very goal-directed verbs (like "complete"). 5 is for non goal-directed verbs (like "rotate").

For brevity, these two continua will be henceforth referred to as generality and goal-directedness.

18 forms were generated (3 rating continua x 2 random orders x 3 sub-lists), and equal numbers of each form were assembled into booklets (1 per booklet) with other questionnaires and distributed to students in an introductory psychology class at Stanford University. Each participant thus one of the 3 possible judgments about 1/3 of one of the word-lists. Students participated in the study in partial fulfillment of a course requirement.

Noun Generality

17 of the noun-generality rating forms were returned, generating ratings based on between 4 and 6 judgments per word. The mean ratings for each word are reproduced below.

| Noun | Generality | Noun | Generality | Noun | Generality |
|---|---|---|---|---|---|
| apron | 1.67 | faucet | 1.83 | pitcher | 2.00 |
| area | 5.00 | fertilizer | 2.33 | place | 4.60 |
| attachment | 4.60 | finger | 2.00 | plant | 3.67 |
| back | 2.33 | floor | 2.17 | plant food | 2.40 |
| bag | 3.50 | food | 4.40 | plate | 1.83 |
| bed | 2.67 | foot | 2.00 | pot | 2.00 |
| bedding | 3.20 | fork | 1.75 | rack | 3.00 |
| blanket | 2.00 | front | 4.33 | rag | 2.40 |
| body | 4.00 | garbage | 3.00 | reed | 2.33 |
| bottom | 2.50 | glass | 3.50 | room | 4.20 |
| box | 3.33 | ground | 4.17 | saxophone | 1.00 |
| buckle | 2.83 | hair | 3.00 | scene | 3.67 |
| cabinet | 2.67 | hand | 1.75 | scoop | 2.33 |
| can | 3.00 | head | 3.33 | screw | 1.83 |
| cap | 3.60 | horn | 2.33 | set | 4.40 |
| cascade | 2.80 | instrument | 4.50 | sheet | 3.67 |
| case | 4.80 | item | 5.00 | shelf | 2.50 |
| chemical | 3.40 | key | 2.60 | side | 4.33 |
| chin | 1.20 | kitchen | 3.40 | silverware | 3.00 |
| cleaner | 3.83 | knife | 2.00 | sink | 2.17 |
| cloth | 3.67 | latch | 2.25 | soap | 1.80 |
| comforter | 1.25 | leaf | 2.17 | solution | 4.20 |
| compartment | 4.00 | ledge | 2.67 | something | 5.00 |
| container | 3.83 | lid | 2.83 | sponge | 1.75 |
| content | 4.67 | linen | 2.80 | spoon | 1.25 |
| corner | 3.20 | lip | 1.40 | spot | 3.80 |
| counter | 2.60 | machine | 4.20 | strap | 3.17 |
| countertop | 1.50 | mattress | 2.00 | string | 2.17 |
| cover | 4.40 | miracle grow | 1.00 | stuff | 4.75 |
| cup | 2.00 | mixture | 4.33 | suitcase | 1.75 |
| cupboard | 1.50 | mouth | 2.00 | table | 2.20 |
| cups | 2.00 | mouthpiece | 1.50 | thing | 5.00 |
| detergent | 2.33 | neck | 2.00 | top | 4.75 |
| dish | 3.00 | neckstrap | 1.83 | trash | 3.00 |
| dishwasher | 1.60 | object | 4.83 | tray | 1.60 |
| door | 2.20 | outside | 4.00 | utensil | 4.17 |
| drain | 1.33 | part | 4.83 | waist | 1.80 |
| drawer | 1.60 | piece | 4.83 | washer | 2.25 |
| edge | 3.60 | pillow | 1.75 | water | 1.50 |
| end | 4.40 | pillowcase | 1.17 | whatever | 5.00 |
| everything | 5.00 | pillowcases | 1.50 | windowsill | 1.33 |

Verb Generality

18 of the verb-generality rating forms were returned, generating ratings based on 6 judgments per word (except for the word "scoop," for which only 3 ratings were obtained due to a typographical error). The mean ratings for each word are reproduced below.

| Verb | Generality | Verb | Generality | Verb | Generality |
|---|---|---|---|---|---|
| add | 3.67 | handle | 3.83 | secure | 3.50 |
| adjust | 4.00 | hold | 3.33 | set | 4.17 |
| approach | 2.67 | hook | 3.00 | shake | 2.83 |
| arrange | 3.67 | insert | 3.00 | shut | 2.67 |
| assemble | 3.83 | kneel | 2.17 | smooth | 2.50 |
| attach | 3.50 | lay | 3.33 | spread | 3.17 |
| bang | 3.17 | lean | 2.50 | stand | 3.67 |
| be | 5.00 | leave | 4.33 | start | 3.67 |
| bend | 3.17 | lick | 2.33 | stick | 3.50 |
| breathe | 3.00 | lift | 3.50 | straighten | 2.83 |
| bring | 3.83 | load | 4.17 | strap | 3.50 |
| change | 4.50 | look | 4.00 | strip | 2.67 |
| check | 4.17 | make | 4.17 | stuff | 4.17 |
| clean | 3.83 | measure | 3.17 | suck | 2.17 |
| clear | 4.00 | mix | 2.67 | take | 4.67 |
| clip | 3.00 | move | 4.33 | throw | 3.33 |
| close | 3.50 | open | 3.17 | tie | 2.83 |
| come | 4.50 | pick | 3.33 | tighten | 2.67 |
| connect | 4.00 | place | 3.83 | toss | 3.67 |
| decide | 4.17 | play | 4.83 | tuck | 3.33 |
| discard | 3.33 | polish | 1.83 | turn | 3.67 |
| do | 5.00 | pour | 2.83 | undo | 4.17 |
| draw | 3.67 | prepare | 3.83 | unfold | 2.67 |
| drop | 3.00 | pull | 3.50 | unhook | 1.33 |
| dump | 3.83 | push | 3.50 | unlatch | 1.83 |
| empty | 3.33 | put | 4.67 | unlock | 2.67 |
| enter | 3.17 | remove | 3.33 | unscrew | 2.67 |
| exit | 3.33 | repeat | 2.83 | unsnap | 2.17 |
| feed | 3.83 | replace | 3.00 | untie | 2.50 |
| fill | 3.67 | return | 3.50 | unzip | 1.40 |
| fit | 4.17 | rinse | 2.33 | walk | 2.83 |
| fix | 4.50 | rotate | 2.50 | wash | 3.83 |
| flatten | 2.67 | rub | 3.00 | water | 2.50 |
| fluff | 1.83 | run | 3.17 | wet | 2.67 |
| fold | 2.17 | scoop | 2.33 | wipe | 2.67 |
| get | 4.50 | scrape | 2.17 | wrap | 3.17 |
| go | 4.17 | screw | 1.50 | | |
| grab | 2.33 | scrub | 1.83 | | |

Verb Goal-directedness

17 of the verb-goal-directedness rating forms were returned, generating ratings based on 3 to 4 judgments per word. The mean ratings for each word are reproduced below.

| Word | Mean | Word | Mean | Word | Mean |
|---|---|---|---|---|---|
| add | 2.25 | handle | 2.75 | secure | 2.00 |
| adjust | 2.00 | hold | 3.20 | set | 2.67 |
| approach | 2.75 | hook | 2.75 | shake | 4.50 |
| arrange | 4.00 | insert | 3.33 | shut | 1.33 |
| assemble | 1.00 | kneel | 2.75 | smooth | 3.00 |
| attach | 2.00 | lay | 3.40 | spread | 3.67 |
| bang | 2.75 | lean | 2.75 | stand | 2.60 |
| be | 4.00 | leave | 3.75 | start | 2.00 |
| bend | 2.50 | lick | 3.20 | stick | 3.00 |
| breathe | 4.00 | lift | 2.50 | straighten | 2.00 |
| bring | 1.50 | load | 2.00 | strap | 2.75 |
| change | 3.00 | look | 2.25 | strip | 2.40 |
| check | 1.75 | make | 1.25 | stuff | 3.00 |
| clean | 2.20 | measure | 1.67 | suck | 4.33 |
| clear | 2.60 | mix | 3.33 | take | 1.25 |
| clip | 3.00 | move | 2.25 | throw | 2.00 |
| close | 1.60 | open | 2.00 | tie | 3.50 |
| come | 1.75 | pick | 2.00 | tighten | 1.67 |
| connect | 3.00 | place | 3.25 | toss | 4.00 |
| decide | 1.00 | play | 3.67 | tuck | 2.20 |
| discard | 2.50 | polish | 2.50 | turn | 3.00 |
| do | 1.25 | pour | 3.00 | undo | 3.25 |
| draw | 2.75 | prepare | 3.50 | unfold | 2.25 |
| drop | 4.33 | pull | 2.00 | unhook | 3.50 |
| dump | 2.50 | push | 1.75 | unlatch | 2.25 |
| empty | 3.20 | put | 2.50 | unlock | 2.00 |
| enter | 3.00 | remove | 2.33 | unscrew | 2.20 |
| exit | 1.75 | repeat | 2.75 | unsnap | 2.67 |
| feed | 2.25 | replace | 2.75 | untie | 2.50 |
| fill | 1.25 | return | 2.25 | unzip | 2.50 |
| fit | 2.50 | rinse | 2.33 | walk | 2.00 |
| fix | 1.00 | rotate | 2.75 | wash | 3.25 |
| flatten | 2.50 | rub | 3.50 | water | 4.25 |
| fluff | 3.33 | run | 2.50 | wet | 4.50 |
| fold | 1.75 | scoop | 2.50 | wipe | 2.25 |
| get | 2.25 | scrape | 3.25 | wrap | 3.00 |
| go | 1.75 | screw | 3.00 | | |
| grab | 3.25 | scrub | 3.25 | | |

Table 1

Correlations between mean ratings for familiarity, frequency of performance, and knowledge of steps of 45 everyday activities.

| | **How Frequent** | **Know Steps** |
|---|---|---|
| **How Familiar** | 0.85 | 0.96 |
| **How Frequent** | | 0.85 |

Table 2

Participants who described activity from memory made longer units than those who described activity on-line. Cell values represent means of mean length per viewing in seconds or mean number of breakpoints per viewing, with SEM in parentheses. (Outliers removed as described in the text.)

| | **Coarse** | **Fine** |
|---|---|---|
| Unit length | | |
| **Describe on-line** | 34.3 (2.61) | 12.8 (1.04) |
| **Describe from memory** | 58.5 (7.41) | 19.6 (2.38) |
| Number of breakpoints | **Coarse** | **Fine** |
| **Describe on-line** | 10.1 (0.92) | 28.9 (2.02) |
| **Describe from memory** | 6.18 (0.36) | 17.4 (0.95) |

Figure 1: Mean ratings of familiarity, frequency of performance, and knowledge of steps for two unfamiliar and two familiar activities.

Figure 2: Schematic representations of the discrete and continuous analyses.

Figure 3: There were substantial differences between participants in overall level of segmentation in both the fine and coarse coding conditions. The top panel is a histogram of mean unit lengths in the coarse coding condition, and the bottom panel is a histogram of mean unit lengths in the coarse coding condition. In both cases, bins are 10 seconds wide, and labeled by their mean.

Figure 4: Agreement of breakpoints across participants. The figure plots the distribution of breakpoint locations for one of the two videotapes of the "making the bed" activity. Time (plotted on the X axis) has been discretized in 4-second bins. The top panel shows the number participants who identified each bin as a breakpoint under coarse-unit coding instructions. The bottom panel shows the number participants who identified each bin as a breakpoint under fine-unit coding instructions. (16 participants in the Describe group watched this videotape, so the maximum possible value on the Y axis is 16.)

Figure 5: One participant's coarse and fine event descriptions for "fertilizing houseplants." Coarse unit descriptions are aligned with the nearest fine unit description based on breakpoint location.

Figure 6: One participant's coarse and fine event descriptions for "making a bed." Coarse unit descriptions are aligned with the nearest fine unit description based on breakpoint location.

Figure 7: Syntactic and semantic features of objects under different description conditions during on-line event segmentation. Error bars represent 95% confidence intervals.

Figure 8: Syntactic and semantic features of verbs under different description conditions during on-line event segmentation. Error bars represent 95% confidence intervals.

Figure 9: Syntactic and semantic features of subjects under different description conditions during on-line event segmentation. Error bars represent 95% confidence intervals.

Figure 10: Effects of training on the magnitude of the alignment effect, as measured by the discrete (top) and continuous (bottom) methods. (Error bars give 95% confidence intervals.)

Figure 11: Effects of group and activity on the magnitude of the alignment effect, as measured by the discrete (top) and continuous (bottom) methods. (Error bars give 95% confidence intervals.)

Figure 12: Syntactic and semantic features of objects under different description conditions in descriptions from memory. Error bars represent 95% confidence intervals.

Figure 13: Syntactic and semantic features of verbs under different description conditions in descriptions from memory. Error bars represent 95% confidence intervals.

Figure 14: Syntactic and semantic features of subjects under different description conditions in descriptions from memory. Error bars represent 95% confidence intervals.

Figure 15: Distribution of distances from grouped-unit breakpoints to the nearest coarse-unit breakpoint.

Figure 1

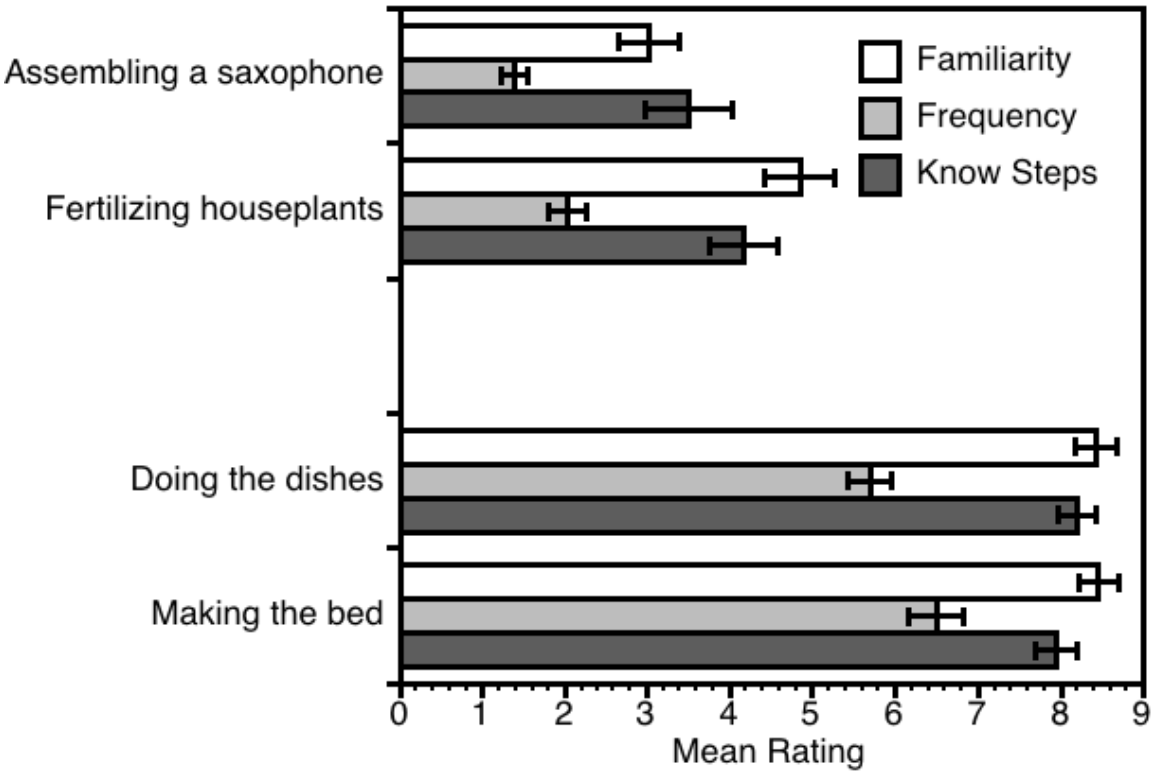

Assembling a saxophone
Fertilizing houseplants
Doing the dishes
Making the bed
Familiarity
Frequency
Know Steps
0
1
2
3
4
5
6
7
8
9
Mean Rating

Figure 2

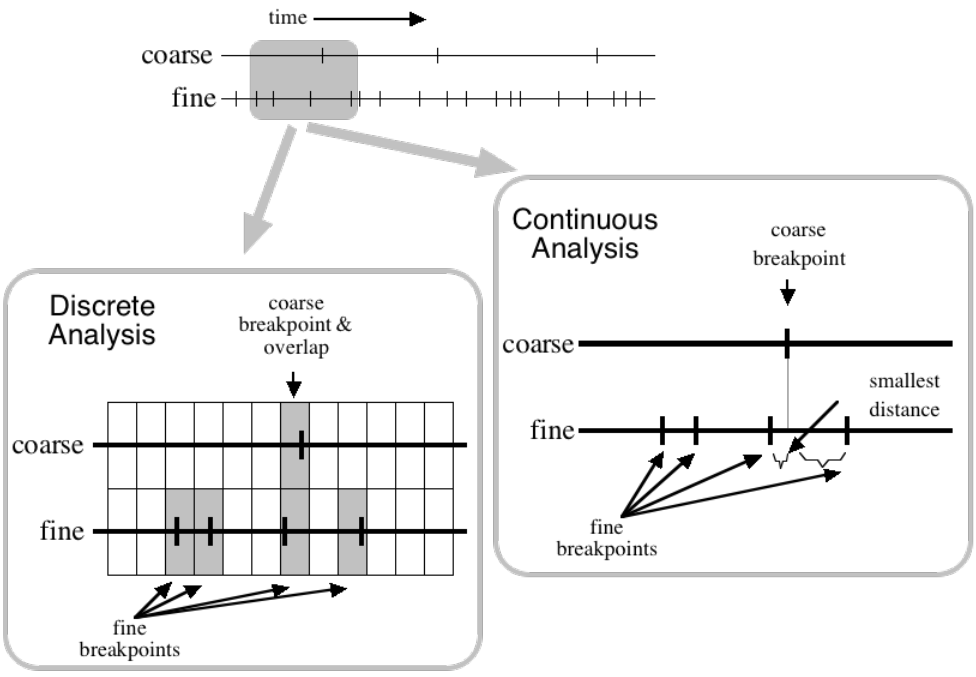

time
coarse
fine
Discrete Analysis
coarse breakpoint & overlap
coarse
fine
fine breakpoints
Continuous Analysis
coarse breakpoint
coarse
smallest distance
fine
fine breakpoints

Figure 3

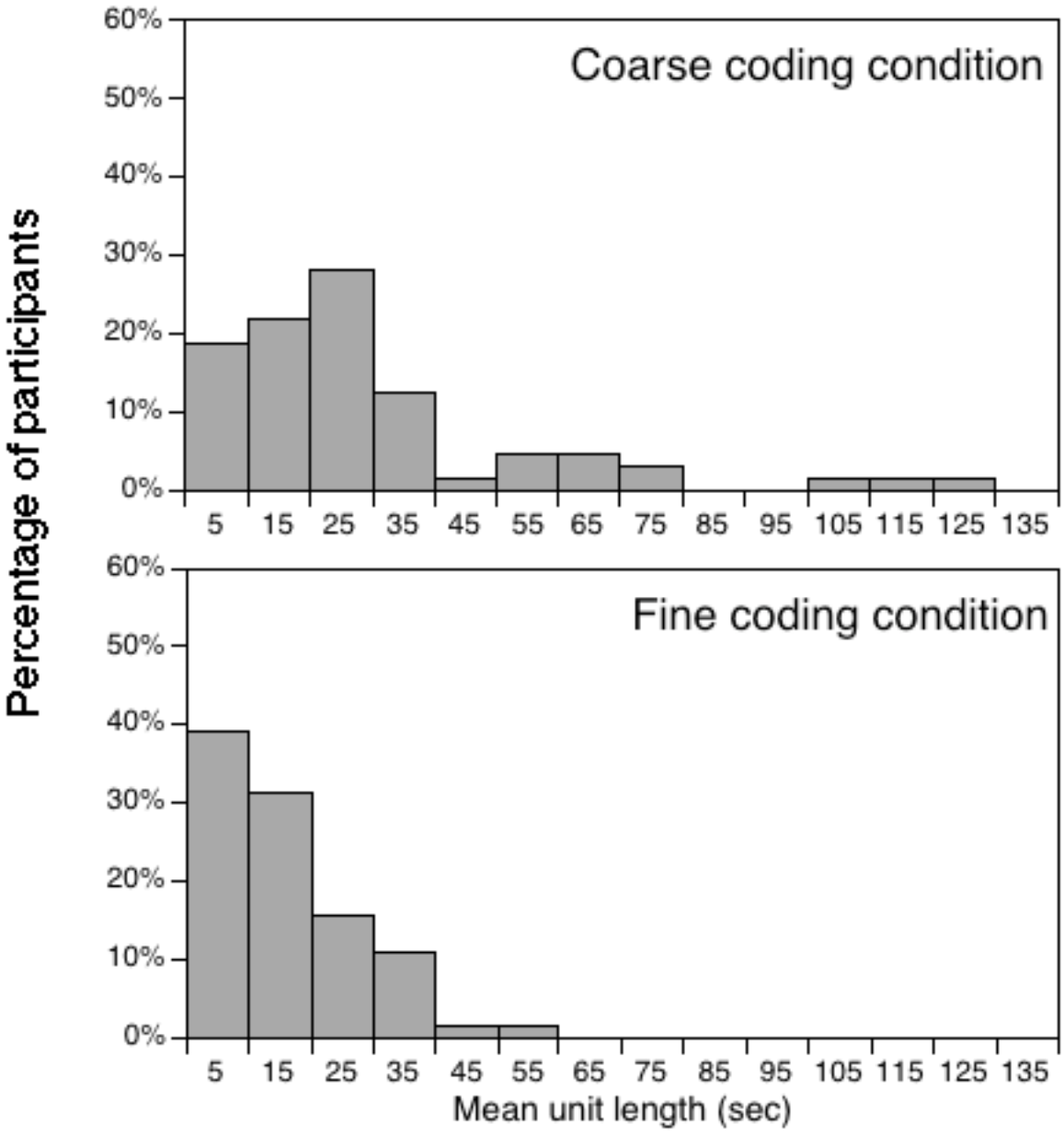

Coarse coding condition
Fine coding condition
Percentage of participants
60%
50%
40%
30%
20%
10%
0%
5
15
25
35
45
55
65
75
85
95
105
115
125
135
Mean unit length (sec)

Figure 4

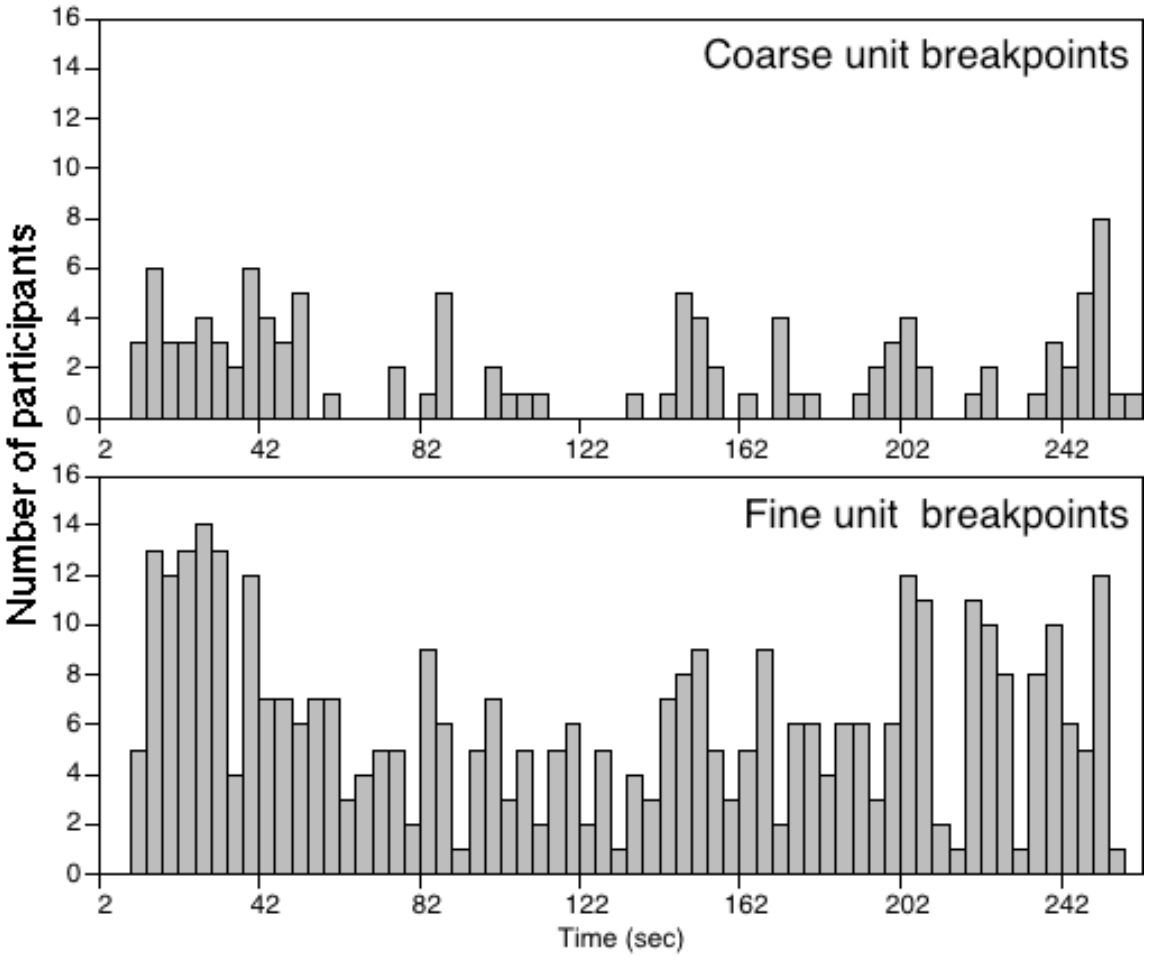

Coarse unit breakpoints
Fine unit breakpoints
Number of participants
Time (sec)
0
2
4
6
8
10
12
14
16
2
42
82
122
162
202
242

Figure 5

| **Coarse unit descriptions** | **Fine unit descriptions** |
|---|---|
| walks in | walks into the room |
| | open door |
| | take out..... food |
| | close door |
| takes the food out | open door |
| | take out pot or thing |
| | put it down |
| | open box |
| | take out something |
| puts the food into the uh.... watering thing | opens bag |
| | takes a scoop |
| | puts it in... thing |
| | puts scoop back |
| | turn faucet |
| adds water | turn off faucet |
| | moves things aside |
| | picks up plant |
| | puts it down |
| starts watering plant | picks up pot.. and waters.. |
| | waters the other side |
| | stops watering |
| | puts pot down |
| | puts plant back |
| | picks up pot |
| | empties it |
| | turn on faucet |
| cleans the watering thing | rinsing |
| | turn off faucet |
| | put down |
| | open door |
| | puts it in |
| | close door |
| | close bag |
| puts food away | close box |
| | open door |
| | put it in |
| | close door |
| that's it , she leaves | walks out |

Figure 6

| **Coarse unit descriptions** | **Fine unit descriptions** |
|---|---|
| walking in | walking in |
| | taking off the blanket |
| | pulling the pillow out of the case |
| | dropping the pillow |
| | pulling pillow out of the case |
| | dropping it |
| | taking out the sheet |
| taking apart the bed | opening a drawer |
| | taking sheets out |
| | unfolding the sheet |
| | putting on the top end of the sheet |
| putting on the sheet | putting on the bottom |
| | unfolding sheet |
| | laying it down |
| | straightening it out |
| | tucking it in |
| | leaning on the bed |
| putting on the other sheet | spreading out the blanket |
| | straightening it |
| | pulling the sheet over the top |
| | straightening it out |
| | lifting the bed up |
| | tucking in the blanket |
| putting on the blanket | picking up pillow, pillowcase |
| | opening it up |
| | putting the pillow in the pillowcase |
| | picking up other pillow |
| | opening the pillowcase |
| | putting the pillow in |
| | putting down the pillow |
| | putting down the other pillow |
| putting in the pillows | walking away |
| walk away | |

Figure 7

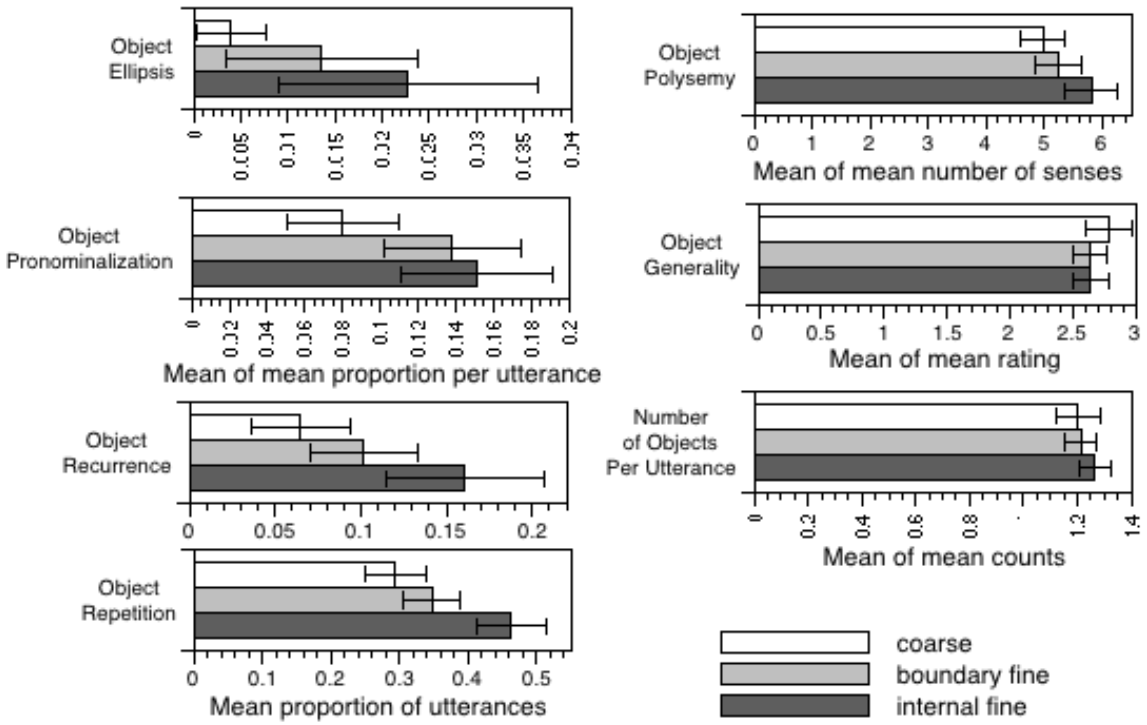

Object Ellipsis
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04
Object Pronominalization
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2
Mean of mean proportion per utterance
Object Recurrence
0
0.05
0.1
0.15
0.2
Object Repetition
0
0.1
0.2
0.3
0.4
0.5
Mean proportion of utterances
Object Polysemy
0
1
2
3
4
5
6
Mean of mean number of senses
Object Generality
0
0.5
1
1.5
2
2.5
3
Mean of mean rating
Number of Objects Per Utterance
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Mean of mean counts
coarse
boundary fine
internal fine

Figure 8

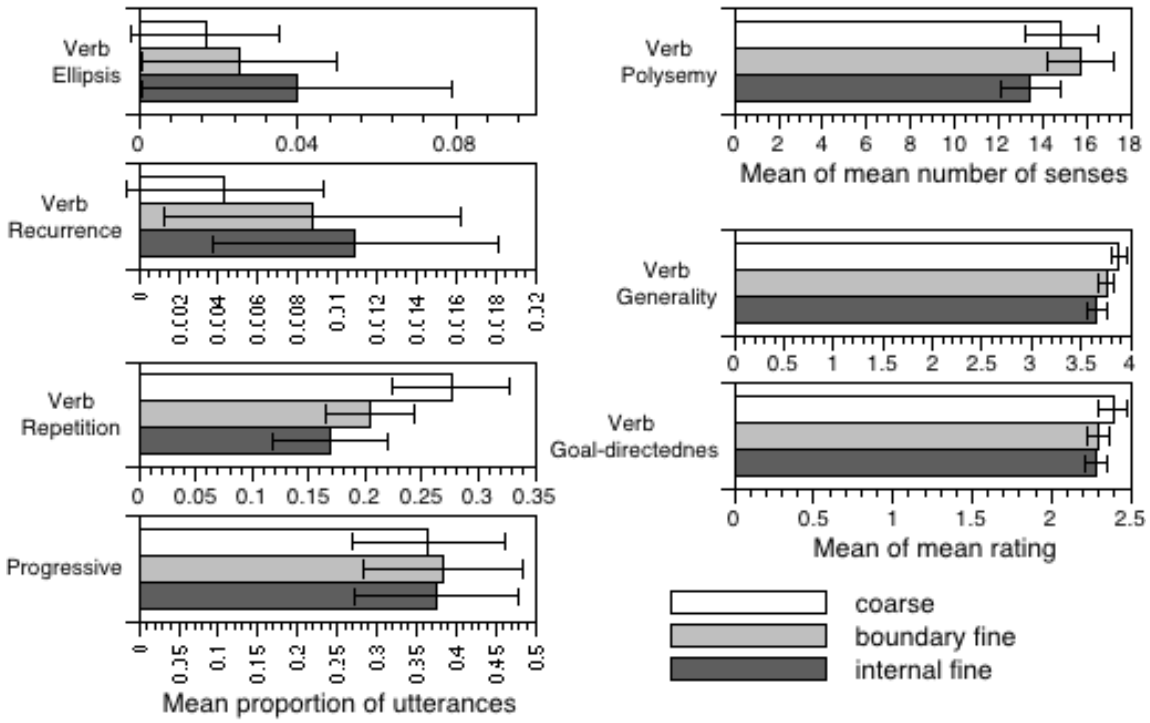

Verb Ellipsis
0
0.04
0.08
Verb Recurrence
0
0.002
0.004
0.006
0.008
0.01
0.012
0.014
0.016
0.018
0.02
Verb Repetition
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Progressive
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
Mean proportion of utterances
Verb Polysemy
0
2
4
6
8
10
12
14
16
18
Mean of mean number of senses
Verb Generality
0
0.5
1
1.5
2
2.5
3
3.5
4
Verb Goal-directednes
0
0.5
1
1.5
2
2.5
Mean of mean rating
coarse
boundary fine
internal fine

Figure 9

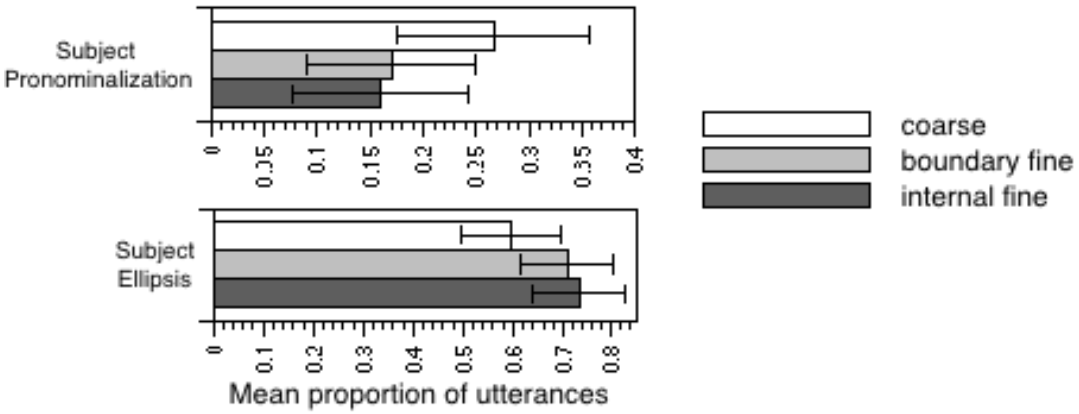

Subject
Pronominalization
0
0.35
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Subject
Ellipsis
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Mean proportion of utterances
coarse
boundary fine
internal fine

Figure 10

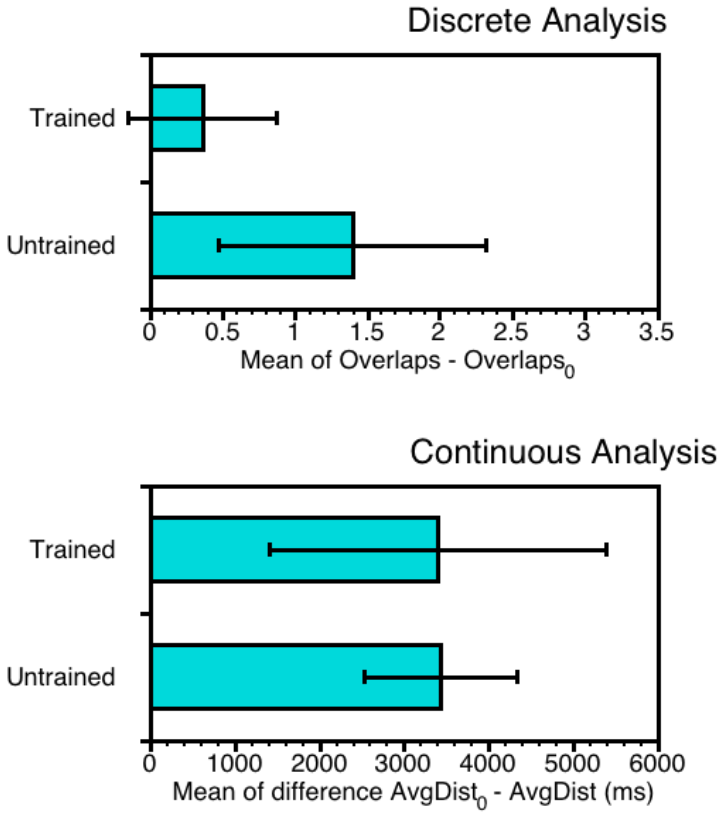

Discrete Analysis
Trained
Untrained
0
0.5
1
1.5
2
2.5
3
3.5
Mean of Overlaps - Overlaps0
Continuous Analysis
Trained
Untrained
0
1000
2000
3000
4000
5000
6000
Mean of difference AvgDist0 - AvgDist (ms)

Figure 11

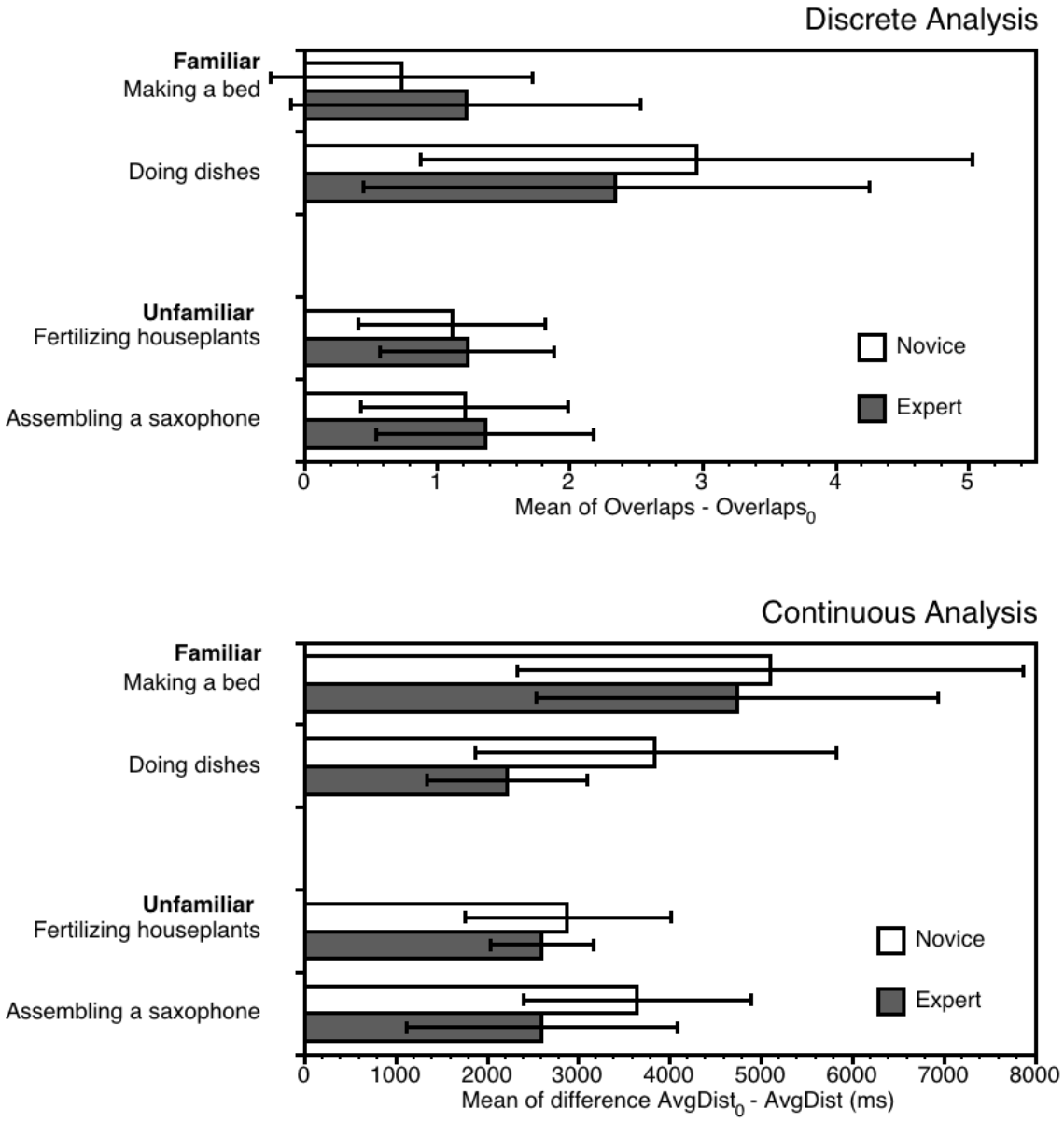

Discrete Analysis
Familiar
Making a bed
Doing dishes
Unfamiliar
Fertilizing houseplants
Assembling a saxophone
Novice
Expert
0
1
2
3
4
5
Mean of Overlaps - Overlaps0
Continuous Analysis
Familiar
Making a bed
Doing dishes
Unfamiliar
Fertilizing houseplants
Assembling a saxophone
Novice
Expert
0
1000
2000
3000
4000
5000
6000
7000
8000
Mean of difference AvgDist0 - AvgDist (ms)

Figure 12

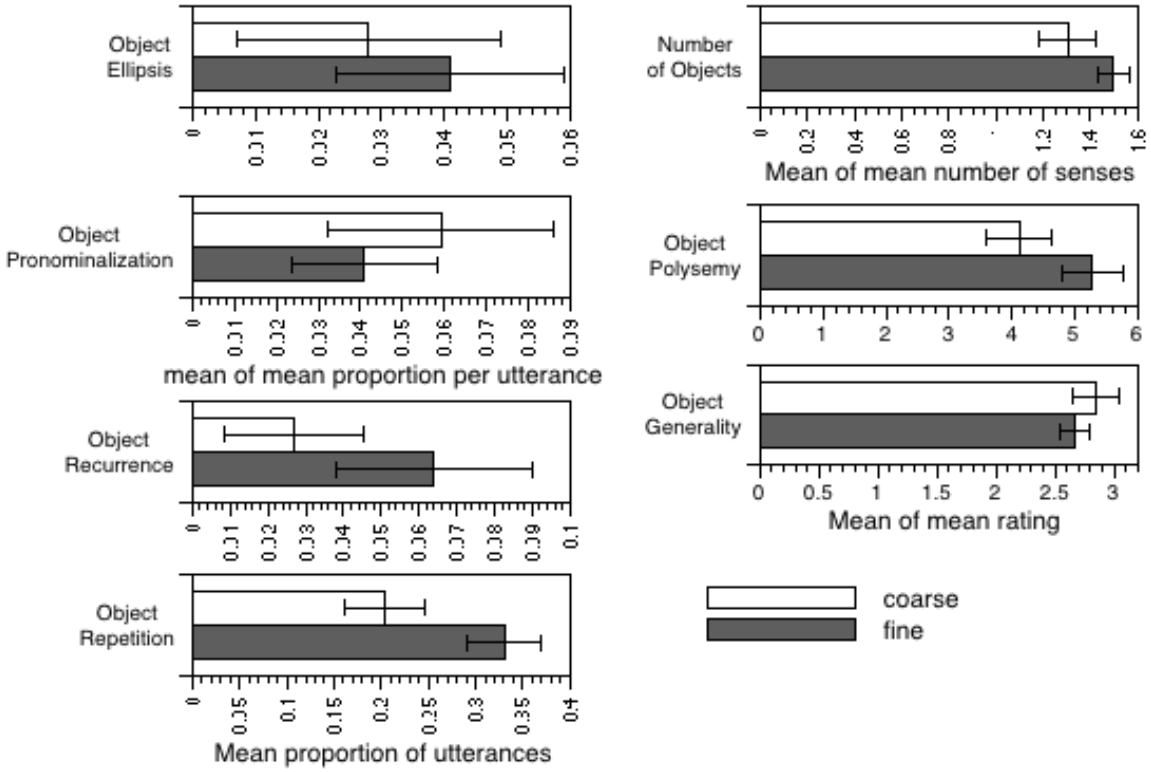

Object Ellipsis
Object Pronominalization
Object Recurrence
Object Repetition
mean of mean proportion per utterance
Mean proportion of utterances
Number of Objects
Mean of mean number of senses
Object Polysemy
Object Generality
Mean of mean rating
coarse
fine

Figure 13

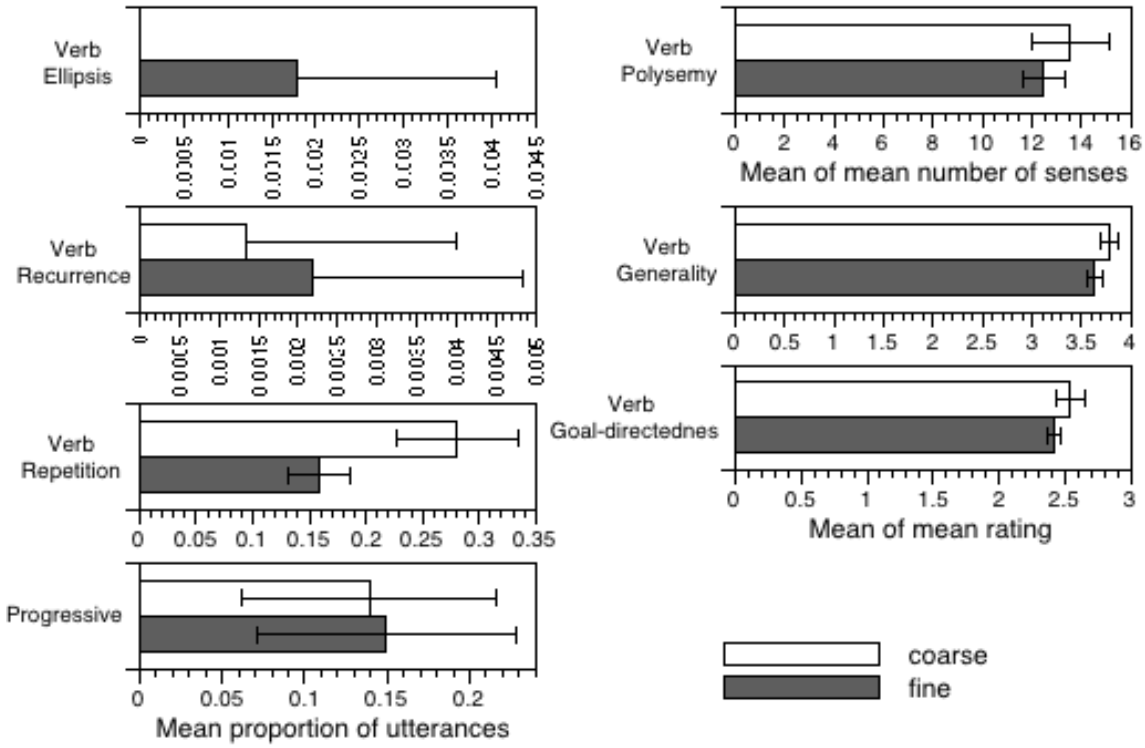

Verb Ellipsis
0
0.0005
0.001
0.0015
0.002
0.0025
0.003
0.0035
0.004
0.0045
Verb Recurrence
0
0.0005
0.001
0.0015
0.002
0.0025
0.003
0.0035
0.004
0.0045
0.005
Verb Repetition
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Progressive
0
0.05
0.1
0.15
0.2
Mean proportion of utterances
Verb Polysemy
0
2
4
6
8
10
12
14
16
Mean of mean number of senses
Verb Generality
0
0.5
1
1.5
2
2.5
3
3.5
4
Verb Goal-directednes
0
0.5
1
1.5
2
2.5
3
Mean of mean rating
coarse
fine

Figure 14

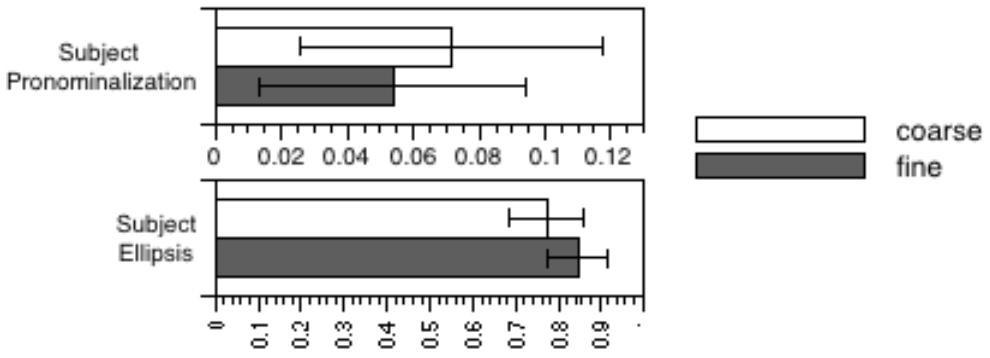
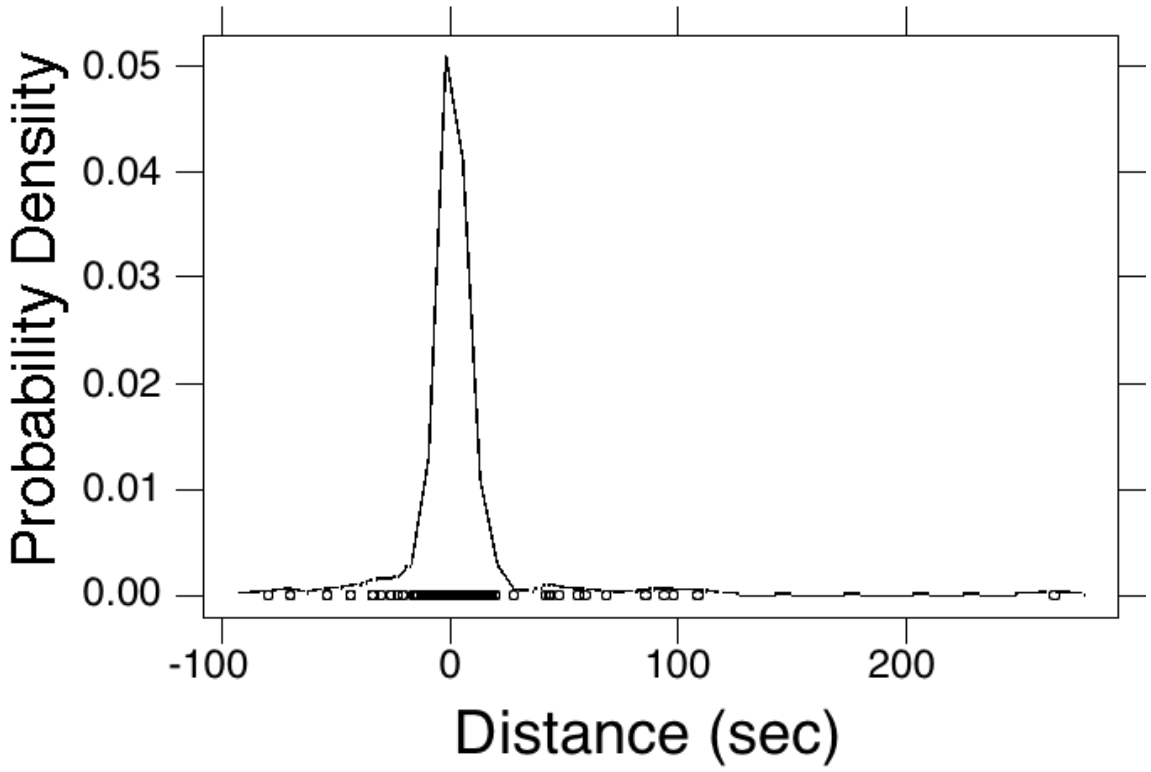

Subject
Pronominalization
0
0.02
0.04
0.06
0.08
0.1
0.12
Subject
Ellipsis
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
coarse
fine

Figure 15


Probability Density
0.05
0.04
0.03
0.02
0.01
0.00
-100
0
100
200
Distance (sec)