Journal of Experimental Psychology: General

The Shape of Action

Bridgette Martin Hard, Gabriel Recchia, and Barbara Tversky

Online First Publication, August 1, 2011. doi: 10.1037/a0024310

Journal of Experimental Psychology: General

0096-3445/11/$12.00 DOI: 10.1037/a0024310

The Shape of Action

Bridgette Martin Hard and Gabriel Recchia
Stanford University

Barbara Tversky
Stanford University and Columbia Teacher's College

How do people understand the everyday, yet intricate, behaviors that unfold around them? In the present research, we explored this by presenting viewers with self-paced slideshows of everyday activities and recording looking times, subjective segmentation (breakpoints) into action units, and slide-to-slide physical change. A detailed comparison of the joint time courses of these variables showed that looking time and physical change were locally maximal at breakpoints and greater for higher level action units than for lower level units. Even when slideshows were scrambled, breakpoints were regarded longer and were more physically different from ordinary moments, showing that breakpoints are distinct even out of context. Breakpoints are bridges: from one action to another, from one level to another, and from perception to conception.

Keywords: event segmentation, action parsing, event memory, goal hierarchies

Understanding others' behavior is crucial in navigating the social world. Whether one is buying a cup of coffee or having a conversation, interpersonal interaction relies on rapidly recognizing actions, interpreting their intentions, anticipating future actions, and coordinating one's own actions accordingly. This ability is remarkable: The quantity and quality of information in human activities is vast and varied. One way the mind copes with such immense information is by reducing and organizing it: The spatial stream of information is segmented into figures, grounds, people, objects, and scenes. Segmenting the temporal stream into meaningful units of action similarly reduces the barrage of information (e.g., Rosch, 1978; Tversky & Hemenway, 1984; Tversky, Zacks, & Hard, 2008). The mind must also organize the units that it segments. Whereas units of people, objects, and scenes have spatial organizations, units of action have temporal ones: parts configured in time. Segmenting and inferring the organization of ongoing action is fundamental for many tasks of life: for perceiving and understanding others' actions, for coordinating our behaviors with others, and for learning and adapting action sequences to suit our own goals. Here, we present a novel approach to the study of action segmentation and organization, investigating the joint time courses of attention, information change, and segmentation.

Segmentation and Organization

Studying how people segment ongoing activities into parts has been brought into the laboratory in a task developed by Newtson (1973). As observers view videos of ongoing activities, they press a key to indicate the moments when one action unit ends and another begins, termed *breakpoints*. Although activities could be segmented at infinite temporal locations, people are remarkably consistent, both with one another and with themselves across viewings, in marking breakpoints (Dickman, 1963; Hard, Tversky, & Lang, 2006; Newtson & Engquist, 1976; Zacks, Tversky, & Iyer, 2001). The action units bookended by these breakpoints are described or recalled with expressions like *rinsed the dish* or *smoothed the sheet*, indicating that they correspond to actions on objects or accomplished goals (e.g., Baldwin & Baird, 1999; Kurby & Zacks, 2008; Newtson, 1973; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zacks, Tversky, & Iyer, 2001).

How are action units organized? Like mental representations of other things in the world, representations of human activities are structured hierarchically into actions and parts of actions, both when performing activities (e.g., Lashley, 1951) and when applying knowledge of them in planning, describing, and recalling. For example, people's listings of features for *going to a restaurant* include components such as *be seated, look at menu,* and *order food*, with subcomponents for each of these (e.g., Bower, Black, & Turner, 1979). The hierarchical organization that emerges when planning and describing activities also emerges when segmenting activities in the laboratory. When participants segment films of activities twice, once into the largest units that seem natural and meaningful (coarse grain) and once into the smallest units (fine grain), large and small units form a hierarchy (e.g., Hard et al., 2006; Kurby & Zacks, 2008; Newtson, 1973; Zacks, Tversky, & Iyer, 2001). Making a bed, for example, is segmented as *putting on*

Bridgette Martin Hard, Department of Psychology, Stanford University; Gabriel Recchia, Symbolic Systems Program, Stanford University; Barbara Tversky, Department of Psychology, Stanford University, and Cognitive Studies Program, Columbia Teacher's College.

Gabriel Recchia is now at the Department of Cognitive Science, Indiana University, Bloomington.

This research was made possible by the financial support of the Norman H. Anderson Research Fund at Stanford University. Preparation of the article was aided by Grants NSF HHC 0905417, IIS-0725223, IIS-0855995, and REC 0440103, the Stanford Regional Visualization and Analysis Center, and Office of Naval Research Grants NOOO14-PP-1-O649, N000140110717, and N000140210534. We gratefully acknowledge Christopher Pratt, Antonietta Moreland, Sanjay Kairam, and Silvia Paz Romero for assistance in the research. Thanks are also due to Gordon Bower, Dare Baldwin, and Jeff Zacks for their helpful advice.

Correspondence concerning this article should be addressed to Bridgette Martin Hard, Department of Psychology, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305. E-mail: bridgette.hard@stanford.edu

the bottom sheet, *putting on the top sheet*, *changing the pillowcases*, and so on. Each of these larger units is segmented into subactions that correspond to subgoals, for *putting on the bottom sheet*: *spreading the sheet, tucking in the corners*, and *smoothing it out*.

Breakpoints Are Privileged

Given that breakpoints separate meaningful units of action, it is intuitive to think of them as conceptually "empty" moments in the flow of action when nothing is happening, like the spaces between words or the pauses between phrases. Yet a wealth of research contradicts this intuition: Breakpoints appear to be especially memorable and meaningful moments in the stream of ongoing action. Single frame images corresponding to breakpoints are better recognized than are images corresponding to within-unit moments (Newtson & Engquist, 1976). Also, sequences of images taken from breakpoints in an action sequence are described more accurately, rated as more intelligible, and ordered more correctly than are sequences of within-unit images (Newtson & Engquist, 1976). What is more, removing breakpoints from ongoing action is disruptive: Films of events with breakpoint moments removed are remembered less well than films of events with within-unit moments removed (Schwan & Garsoffky, 2004).

Breakpoints thus seem less like the spaces between words and more like the inflection points in an object contour that define the object's parts. Like inflection points, breakpoints correspond to discontinuities in the flow of information. A number of studies have shown, for example, that physical movement increases at breakpoints. In one study, the number of joints moving on an actor was greater for breakpoints than for within-unit points (Newtson, Engquist, & Bois, 1977). In other research, observers watched and segmented films of animated geometric figures that could be interpreted as intentional movements of agents. In one study, the number of different kinds of motion changes, such as changes in speed or direction was greater for breakpoints than for within-unit moments (Hard et al., 2006). In another study, regression analyses showed that breakpoint locations were well predicted by low-level physical changes (Zacks, 2004). In more recent work, researchers found more movements of actors' hands at breakpoints, in activities like assembling Lego blocks or folding laundry (Zacks, Kumar, Abrams, & Mehta, 2009). Together, these studies suggest that there is greater change in the action stream at breakpoints and, consequently, greater amounts of new information to process.

That action change is greater at breakpoints than at ordinary moments in the behavior stream may have important consequences for how people process activities in real time. According to event segmentation theory (EST; Kurby & Zacks, 2008; Reynolds, Zacks, & Braver, 2007; Zacks et al., 2007), observers understand activities by constructing a set of ongoing representations of "what is happening now," called *event models*. Event models are generated using both bottom-up input from the action stream and top-down input from knowledge structures in memory called *event schemata.* Event models serve to generate perceptual predictions that guide processing of the activity. Within action units, event models make good predictions, meaning that the actor's movements continue to match the observer's representation of "what is happening now." But as actions are completed and the situation changes, prediction error increases, and the contents of the event models become unstable. When prediction error is sufficiently high, breakpoints are perceived, and the event models must be updated. Updating the event models requires increased processing of bottom-up input and activating appropriate event schemata to replace the previous ones.

According to EST, breakpoints involve a convergence of major changes in bottom-up perceptual information and in top-down conceptual information. This view of breakpoints is consistent with available data. Studies performed outside the laboratory show that breakpoints are marked at conceptual changes in setting, actor, object, or action (Barker & Wright, 1955). In the laboratory, the play-by-play descriptions of actions indicate that breakpoints also coincide with changes in intentions as they mark the accomplishment of goals or subgoals (Zacks, Tversky, & Iyer, 2001). The convergence of perceptual and conceptual changes at breakpoints has parallels in the domain of physical object segmentation. For objects, the convergence of perceptual and conceptual information allows people to segment objects on the basis of perceptual features and to bootstrap those segment boundaries to inferences about more abstract behavioral and functional information (Tversky & Hemenway, 1984). The convergence of perceptual and conceptual information in action segmentation may allow similar bootstrapping.

Breakpoints are transitions from action to action, separating event parts as well as bridging them. Because a new action is often begun as the current one is completed (Mennie, Hayhoe, & Sullivan, 2007), breakpoints often simultaneously capture the completion of one action and the initiation of the next. Because so much happens at breakpoints of both a perceptual and a conceptual nature, breakpoints should be moments not only of greater physical change but also of heightened attention, that is, increased intensity of information processing. This analysis can account for many of the previous findings. If breakpoints are accompanied by greater information and heightened attention, then they should be especially memorable and comprehensible (Newtson & Engquist, 1976), and removing them from an action sequence should be especially disruptive (Schwan & Garsoffky, 2004).

Predictions and Approach

According to the present analysis, breakpoints are privileged because they are inherently more informative than are other moments in the action stream: They correspond to significant perceptual and conceptual transitions. This reasoning leads to a set of predictions, tested in the first experiment. Because of their greater informativeness, breakpoints should demand heightened attention and processing. This prediction is tested with a new measure of action processing: *dwell time*. Dwell time is the amount of time observers look at still-frame images—slides—taken at equal intervals across an action sequence. Dwell time should be longer at breakpoints. Second, breakpoints should correspond to greater change in the action stream. This prediction is tested with a new, objective measure of physical *change* from slide to slide. Together, these new measures, dwell time and change, provide a detailed picture of the joint time courses of attention and information change, with respect to breakpoints, as activities unfold. Next, if breakpoints are indeed the informative moments of an action stream that define its contour or "shape," then greater attention to breakpoints should enhance overall memory for the action sequence. The first experiment examines these predictions.

In the second experiment, we investigate a corollary of the analysis. Because they mark major perceptual and conceptual changes as well as transitions between actions, breakpoints may be informative, even out of context. The second experiment tested this possibility by examining dwell time and change across chronologically ordered versus scrambled sequences of slides. If breakpoints are informative even out of context, then we expect them to draw longer dwell time, even when temporal sequencing is disrupted.

Experiment 1: Looking Time and Change Are Maximal at Breakpoints

In this experiment, we examined the joint time courses of attention, physical change, and breakpoints using a novel dwell time method. Films of action sequences were sampled every second and turned into slideshows. Observers watched a slideshow at their own pace, under instructions to study the action sequence for later recall; the variable of interest was dwell time (i.e., time spent looking at each slide). When there is more information to process, presumably at breakpoints, dwell time should increase. Breakpoints were determined by later segmentation of the source films into units at multiple levels of granularity. A second novel method provided an index of information change. The pixel-to-pixel change from slide to slide was computed after the slides had been put through a filter that extracted figures and eliminated possible confounding effects such as lighting changes. The simple prediction is that dwell time and change should increase at breakpoints.

The dwell time procedure can add substantially to earlier evidence that observers spontaneously segment events hierarchically. Previous research demonstrating hierarchical segmentation involved explicitly instructing participants to divide activities into coarse or fine units on different viewings and then showing that the boundaries of coarse and fine units temporally align (Zacks, Tversky, & Iyer, 2001). Although participants were not instructed to select coarse and fine units that were hierarchically related, simply instructing participants to segment activities on different levels may bias them toward a hierarchical interpretation of activities. The dwell time procedure, because it measures attention in the absence of any specific instructions to segment the activity, presents a more naturalistic portrait of action processing. If observers look longer at both coarse and fine breakpoints, then that implies that they are segmenting at both levels simultaneously and spontaneously. Such a finding would be consistent with brain activation data, which show that even under passive viewing of an activity, several brain regions respond selectively to breakpoints at both coarse and fine levels of segmentation (Zacks, Braver, et al., 2001). Moreover, dwell time might vary systematically with level of segmentation. Because coarse breakpoints mark larger changes in goals of actors (Zacks, Tversky, & Iyer, 2001) than do finer breakpoints, they may correspond to larger amounts of physical change as well. Larger conceptual and physical changes at coarse breakpoints should elicit longer dwell time.

After viewing the slideshow, participants recalled as many actions from the slideshows as they could. If breakpoint moments carry more information, then differential looking at breakpoints should boost recall. What might we predict about the relation between recall and differential looking at fine versus coarse breakpoints? Previous evidence with a segmentation task indicates that asking observers to segment activity at a fine level instead of a coarse level boosts how much detail they can later recall (Hanson & Hirst, 1989). In that task, viewers were instructed to attend to only one level of segmentation, coarse or fine, for the purpose of the segmentation task. A different outcome might emerge in the dwell time procedure, in which observers are free to attend to whatever they like. It is well known that organization improves memory (e.g., Tulving, 1962). Thus, modulated attention that reflects the hierarchical structure of the activities should further enhance organization and sense making. Longer looking at coarse breakpoints, specifically, because they correspond to changes in the overarching goals of the activity, may suggest that observers are encoding the hierarchical structure of an activity. Thus, if observers are attending to multiple levels of abstraction simultaneously, fine units as well as coarse, then greater looking time at coarse breakpoints should best predict recall.

Furthermore, if attention to coarse breakpoints is associated with perceiving the hierarchical structure of the activity, then it should predict how hierarchically organized a participant's pattern of segmentation is during the later segmentation task. We explored two approaches to measuring degree of hierarchical organization in segmentation patterns: a quantitative approach based on the approach of Zacks, Tversky, and Iyer (2001) and a new qualitative approach.

Method

Stimuli. Four familiar, potentially hierarchically organized activities were selected: cleaning a dorm room (*cleaning*), eating breakfast (*breakfast*), putting on makeup (*makeup*), and assembling a television cart (*TV cart*). Each of these activities was performed, unscripted, by an actor and was filmed where the actor typically performed the activity. The films ranged in length from 156 s to 247 s. For the films cleaning, breakfast, and makeup, an actor was simply asked to perform the activity while being filmed. For the film TV cart, the actor was a participant in a previous experiment who had given permission to use his film in future research.

Each film was shot with a stationary digital film camera at a rate of 30 frames per second. A slideshow was created of each film by selecting a single frame from the middle of each 1-s interval (frame 15, 45, 75, etc.), resulting in slideshows ranging in length from 156 to 247 slides. A sampling rate of 1 slide per second was selected because it simplified the comparison of dwell time to segmentation. Pilot testing indicated that the slideshows were comprehensible at this rate and easy to view in a short sitting.

Participants and design. The experimental design was a $4 \times 2 \times 3$ mixed factorial design, with activity (cleaning, breakfast, makeup, and TV cart) and segmentation order (coarse to fine, fine to coarse) varied between-subjects and segmentation level (fine, intermediate, coarse) varied within. Forty Stanford undergraduates participated in exchange for pay or course credit.

Dwell time procedure. Participants were instructed that they would view a single slideshow on the computer; 10 participants viewed each activity. They were asked to pay careful attention to the actions being performed because they would recall those actions later. The slideshow was self-paced; participants were told to spend as much time looking at each slide as they wished and to advance to the next slide by pressing the spacebar. The procedure did not permit participants to return to slides they had already viewed. The experimenter left the room during slideshow viewing. Afterward, participants were given a sheet of paper on which to list as many actions as they could remember as accurately as possible. Participants were not told how much detail to recall, and if they asked, they were told to do whatever seemed reasonable to them. The recall task was not timed.

Segmentation task procedure. After recalling the actions in the slideshow, participants viewed and segmented a film version of the activity that they had just viewed as a slideshow. Because we were interested in whether dwell times would be modulated not only by the breakpoints but also by the hierarchical structure of the activities, it was critical to determine how participants would segment the activities when attempting to organize them hierarchically. Thus, before segmenting the films, participants were given specific examples of how activities can be broken down into hierarchically organized parts to bias them toward looking for hierarchical structure in the segmentation task.

Previous studies have examined segmentation at only two levels, coarse and fine (Hard et al., 2006; Newtson, 1973; Zacks, Tversky, & Iyer, 2001). Because behavior is likely to be understood on more than two levels, participants in the present study identified units on at least three levels. On the first viewing, participants in the coarse to fine condition were asked to divide the activity into the largest units that seemed natural and meaningful. On the second viewing, they were asked to divide the activity into smaller units and, on the third viewing, into even smaller units. Participants in the fine to coarse condition segmented three times in the opposite order. All participants viewed and segmented on at least three levels and were allowed to continue a fourth and even a fifth time.[1]

Calculation of physical change. In order to provide a metric of low-level slide-to-slide change, we adapted change-detection algorithms common in computer science (e.g., Radke, Andra, Al-Kofahi, & Royam, 2005). Because we wanted an index reflecting change in human action, the slides were preprocessed through an edge-detection filter to remove possible confounds due to variations in color, lighting, and the like. This left an array of pixels varying in one dimension, brightness, for each slide. Then, the difference between corresponding pixels of adjacent slides was computed, yielding a change value for each slide. A more detailed description of the metric is described in Appendix A. The only activity in the films was the movements of a single actor, and the films were taken from a camera on a tripod at a single visual angle, so that any changes in pixels were attributable to movements and reorientations of the actor's body and any objects on which he or she acted. When the actor moved or reoriented dramatically from one slide to the next, then the brightness of the corresponding pixels in the two slides differed more than when there was little movement or reorientation.

Results

For all analyses, the criterion for significance was an alpha level of .05. The p values for insignificant effects are reported only when F or t values are greater than 1. For significant analysis of variance (ANOVA) effects, we report partial eta squared (η_p^2) as an estimate of effect size. For significant t test effects, we report Cohen's d. When necessary, a Greenhouse-Geisser correction was applied to the degrees of freedom to correct for violation of sphericity.

Looking time data. For each participant, the first slide (and corresponding 1-s bin of the segmentation data) was excluded from all analyses because data collected during the first 1-s interval are atypical because participants are adjusting to the task. Looking times were then transformed with a $\log_{10}$ function to reduce positive skewness. There was a trend for participants to look longer at slides presented earlier in the slideshow, suggesting that it takes some time for viewers to make general sense of the activity. This trend was well described by a power function (see Figure 1). Fitting power functions to the looking time data for each participant accounted for an average 48% of variance ($SEM = 3\%$) in looking time. The looking time data were detrended to factor out the influence of longer early looking times. A power function was fit to the looking time data for each participant with a curve estimation regression technique, and all further analyses were conducted on the residuals from this regression analysis.

Segmentation data. There were no main effects of activity (cleaning, breakfast, makeup, or TV cart) or segmentation order (coarse to fine, fine to coarse) on the mean number of breakpoints that participants identified, $F(1, 32) < 1$, $MSE = 692.76$, for both, with no significant interactions. On average, fine breakpoints were marked every 6.8 s, intermediate breakpoints were marked every 16.0 s, and coarse breakpoints were marked every 42.0 s.

Recall that participants were permitted to segment the activity more than three times if they wanted to. Sixteen participants segmented the activity a fourth time (eight participants in each segmentation order condition). There were no significant differences in the mean number of fine, intermediate, and coarse breakpoints (based on the first three segmentations each participant performed) between these participants and those who segmented only three times, $ts(38) < 1$. For the eight participants who segmented a level coarser in the fine to coarse condition, their fourth segmentation was on average more coarse (M length = 97.8 s) than the average coarse segmentation for the 12 participants who only segmented three times (M length = 46.4 s). Similarly, for the eight participants who segmented a level finer in the coarse to fine condition, their fourth segmentation was on average finer (M length = 4.6 s) than the fine segmentation of the 12 participants who segmented only three times (M length = 7.6 s). Thus, those who segmented four times added a level of abstraction, compared with those who segmented three times.

Does looking time increase at breakpoints? If breakpoints are more informative than other moments, observers should look longer at breakpoint slides than at slides within units. Before testing this prediction, slides needed to be categorized as falling at or between unit boundaries. Remember that the slides in the slideshow were taken from the midpoint of each 1-s interval of the film. Thus, for each participant, the looking time for each slide was binned as a unit breakpoint if the participant marked a breakpoint

[1] We expected that some viewers would segment the activities on more than three levels of abstraction, and we wanted to have those levels of segmentation available for analysis in the event that the first three levels of segmentation did not reasonably align across participants.

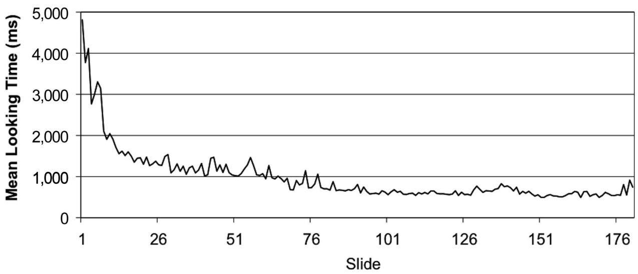

Figure 1. Mean looking time in ms, for the cleaning slideshow.

within the corresponding 1-s interval of the later segmentation task; a slide was binned as within unit if the participant did not mark a breakpoint within that interval. The means and standard deviations were then calculated for each bin. Because there were unequal numbers of slides in the two bins, the mean looking time for each bin was divided by its standard deviation. The resulting looking time score is essentially an effect size estimate, representing how much the mean looking time in a given bin differed from the average looking time for all slides.[2] Because these looking time scores are based on the detrended residual looking time values, they can be negative.

Because there were no effects or interactions of activity or segmentation order, the data were collapsed across conditions. As predicted, participants looked longer at breakpoint slides (M = 0.18, SEM = 0.03) than within-unit slides (M = −0.02, SEM = 0.01), paired $t(39) = 5.62$, $d = 1.48$. This supports the hypothesis that breakpoints elicit increased attention.

Participants looked longer at breakpoints, but did the amount of looking vary as a function of segmentation level? Looking times already binned as unit breakpoints were further binned as fine, intermediate, or coarse breakpoints. Because people identified coarse units that were composed of intermediate units and intermediate units that were composed of fine units, there was often overlap among breakpoints. For example, a slide corresponding to a coarse breakpoint was also likely to correspond to both an intermediate breakpoint and a fine breakpoint. To ensure independence of the bins, looking times were categorized at the highest level of organization. For example, if a looking time corresponded to both coarse and intermediate breakpoints, it was binned only as a coarse breakpoint. This procedure did not change the overall pattern or reliability of the results reported here.

Because there were no effects or interactions of activity or segmentation order, the data were collapsed across conditions. A repeated-measures ANOVA with segmentation level (fine, intermediate, coarse) as a factor revealed that looking time varied significantly with segmentation level, $F(1.7, 64.8) = 5.43$, MSE = 0.13, $\eta_p^2 = .12$. This effect was characterized by a linear trend: Looking times were longest for coarse breakpoints (M = 0.35, SEM = 0.09), followed by intermediate breakpoints (M = 0.22, SEM = 0.05) and then fine breakpoints (M = 0.11, SEM = 0.03), $F(1, 39) = 8.23$, MSE = .14, $\eta_p^2 = .17$.

Does physical change increase with breakpoints? A simple explanation for participants looking longer at breakpoints than at within-unit slides would be that breakpoints looked somehow aberrant (e.g., blurry or containing unnatural-looking body postures). However, as Figure 2 shows, visual inspection of the slides reveals no obvious, peculiar characteristic of breakpoint slides that would draw attention, compared with within-unit slides. So what sets breakpoint moments apart from other moments? In the next analysis, we explored whether breakpoint slides corresponded to heightened amounts of relative physical change, compared with ordinary moments. The makeup slideshow was excluded from this analysis because it was determined (after data had been collected) that many of the slides had digital "scratches"—irregularities in small groups of pixels on the screen. The scratches were an unintended side-effect of encoding the digital video at too high a compression level and introduced unacceptable levels of noise into the change calculations.

For the remaining three activities, the change values were first compared across activities for differences. The activities did differ on the change measure, $F(2, 651) = 253.85$, $MSE = 9.61 \times 10^{11}$, $\eta_p^2 = .44$, with the highest values in cleaning ($M = 6.32 \times 10^6$, $SEM = 1.07 \times 10^5$), followed by breakfast ($M = 5.32 \times 10^6$, $SEM = 5.52 \times 10^4$) and then TV cart ($M = 4.17 \times 10^6$, $SEM = 3.96 \times 10^4$). This makes sense because in cleaning, the actor moved about a room making many different whole body movements; in TV cart, the actor was stationary at a table, moving parts with his arms. Activity did not interact with any of the effects reported next, so the change values were standardized to equate the activities and for ease of interpretation.

The standardized change values for each slide were categorized as falling at breakpoints or within-units using the same binning procedure used for looking times. Change values binned as breakpoints were further binned as coarse, intermediate, or fine breakpoints. The mean change value for each bin was then adjusted by dividing it by its standard deviation. The resulting number is the final change index for each bin.[3] Breakpoint slides corresponded to greater amounts of change (M = 0.24, SEM = 0.04) than within-unit slides (M = −0.14, SEM = 0.03), paired $t(29) = 7.25$, d = 2.08. Because the change index was based on standardized data, the average change per slide was 0. This means that change at breakpoints was higher than average and that change within-units was lower than average. Figure 3 shows an example from the cleaning slideshow. The finding of greater physical change at breakpoints substantiates previous work with other stimuli and other measures of change (e.g., Hard et al., 2006).

Change indices also varied significantly as a function of segmentation level, $F(1.24, 36.06) = 9.28$, MSE = 3.89, $\eta_p^2 = .24$. There was a significant linear trend, such that coarse breakpoints had the highest change index (M = 0.71, SEM = 0.16), followed by intermediate breakpoints (M = 0.25, SEM = 0.05) and then fine breakpoints (M = 0.19, SEM = 0.04), $F(1, 29) = 11.38$, MSE = 4.03, $\eta_p^2 = .28$.

The joint time course of looking and physical change. Until now, the analyses have compared breakpoints with dwell time

[2] The reported differences in looking time between breakpoints and within units, and among breakpoints at different levels, were significant even if the means for each bin were not divided by their standard deviations.

[3] The reported differences in physical change between breakpoints and within units, and among breakpoints at different levels, were significant even if the means for each bin were not divided by their standard deviations.

Figure 2. Example slides from each slideshow (from top to bottom: cleaning, breakfast, makeup, and TV cart). On the left are slides that had a high likelihood of being binned as coarse breakpoints in Experiment 1. On the right are slides that were consistently binned as within-unit slides. Black boxes over the actors' faces are used here and in later figures to conceal the actors' identities. These boxes are used solely for the purposes of this publication and were not present in the original slideshow shown to participants.

and with change. Now we examine the time course of looking and physical change before and after breakpoints at various levels. We analyzed a 6-s window of time surrounding breakpoints for each participant.[4]

For each participant, we created new bins for the looking times and change values. These bins corresponded to the three slides preceding a coarse breakpoint slide and the three slides following it. Bins were also created for slides surrounding intermediate and fine breakpoint slides. For each participant, looking time and physical change measures were calculated for each bin by determining the average looking time and change value for all slides in that bin and dividing those averages by the standard deviations for that bin to obtain a looking time score and change index.

The results are shown in Figure 4. For coarse breakpoints, looking time gradually increased several slides prior to a breakpoint, peaked at the breakpoint, and then tapered off for several slides thereafter. This pattern was reliably characterized by a quadratic function, $F(1, 39) = 6.38$, $MSE = 0.28$, $\eta_p^2 = .14$. This quadratic pattern was also reliable for intermediate, $F(1, 39) = 5.15$, $MSE = 0.14$, $\eta_p^2 = .12$, and fine breakpoints, $F(1, 39) = 5.38$, $MSE = 0.05$, $\eta_p^2 = .12$. The temporal course of physical change was globally similar to the temporal course of looking time. Although the peak of physical change was at the coarse breakpoint, change started to be visibly greater in the slide preceding the breakpoint and tapered off for several slides thereafter. Similarly, this pattern was reliably characterized by a quadratic function, $F(1, 29) = 14.14$, $MSE = 0.48$, $\eta_p^2 = .33$. This quadratic pattern was also reliable for intermediate, $F(1, 29) = 12.71$, $MSE = 0.04$, $\eta_p^2 = .31$, and fine breakpoints, $F(1, 29) = 6.34$, $MSE = 0.04$, $\eta_p^2 = .18$. For both looking time and physical change, the relative differences between coarse, intermediate, and fine levels of segmentation appeared to be highly consistent in the window of time surrounding a breakpoint.

Despite their striking similarities, the curves for looking time and physical change diverge slightly before breakpoints. Both looking time and the change index increase prior to breakpoints, but looking time appears to increase several slides prior to increases in physical change. Figure 4 suggests that looking time started to increase three slides (seconds) prior to a breakpoint, whereas physical change started to increase only one slide before the breakpoint.

Relating change to looking time. The previous analyses showed that physical changes in the behavior stream, like looking time, increased at breakpoints and increased more at coarser levels of segmentation. The previous analyses also showed similar time courses for looking time and physical change. Indeed, the change index for each slide reliably correlated with looking time for that slide for 20 of the 30 participants. On average, the correlations for each participant between looking time and change were not large ($M = .19$, $SEM = .04$) but were significantly different from 0, one sample $t = 4.77$. This means that people generally looked longer at slides that reflected more change. The low correlations could, among other possible causes, be due to the fact that looking time started to increase quite a bit prior to breakpoints and prior to increases in the change measure. This discrepancy prompted further analyses: To what extent does change account for looking time?

A series of regression analyses were performed to find out. For each slide in the slideshows breakfast, cleaning, and TV cart, three values were calculated. First, the mean looking time was determined by pooling and averaging the looking times of the 10 participants who viewed that slide. These mean looking times were then transformed with a $\log_{10}$ function to reduce positive skewness, and the collection of mean looking times for each slideshow were detrended by fitting a power function and extracting the residuals. Second, the overall likelihood that a given slide was selected as a breakpoint was calculated. For each participant, the number of breakpoints corresponding to a given slide was divided by that participant's total number of breakpoints. These values were summed across all participants who viewed that slide to yield the total likelihood that a slide was selected as a breakpoint. The same process was used to determine the likelihood of being se-

[4] This analysis was performed separately for coarse, intermediate, and fine breakpoints. As for the previous analyses of looking time and physical change, we analyzed only intermediate breakpoints that did not overlap with coarse breakpoints and only fine breakpoints that did not overlap with either coarse or intermediate breakpoints.

Figure 3. Because breakpoints involve reorientations of the actor's body as she shifts from one goal to the next, the difference between a breakpoint slide (top right) and its immediately preceding slide (top left) is greater than the difference between two adjacent within-unit slides (bottom).

lected as a fine, intermediate, or coarse breakpoint. Third, the change value for that slide relative to the previous slide was calculated. All calculations just described were then standardized for each film, and submitted to regression analysis.

A first series of regression analyses addressed whether physical change accounts for why observers look longer at breakpoints overall. First, we confirmed that at the group level, the likelihood that a slide was selected as a breakpoint positively predicted the mean looking time for that slide (R^2 = .08, β = .29), $t(1)$ = 7.73. The change value for a given slide also positively predicted mean looking time (R^2 = .11), β = .33), $t(1)$ = 9.06, as well as the likelihood that a slide was selected as a breakpoint (R^2 = .11, β = .33), $t(1)$ = 8.91. Given the relation among these three variables, the next question was whether change accounts for why observers look longer at breakpoints. If so, then the likelihood that a slide was selected as a breakpoint should not predict mean looking time when change is included in the analysis as an independent variable. This was not the case; even controlling for change, the likelihood that a slide was selected as a breakpoint significantly predicted looking time (β = .20), $t(1)$ = 5.25, and explained significant additional variance (change R^2 = .04), $F(1, 651)$ = 27.61. This implies that quantity of physical change, as measured by our crude technique, is a significant contributor to looking time at breakpoints but cannot account for all of it.

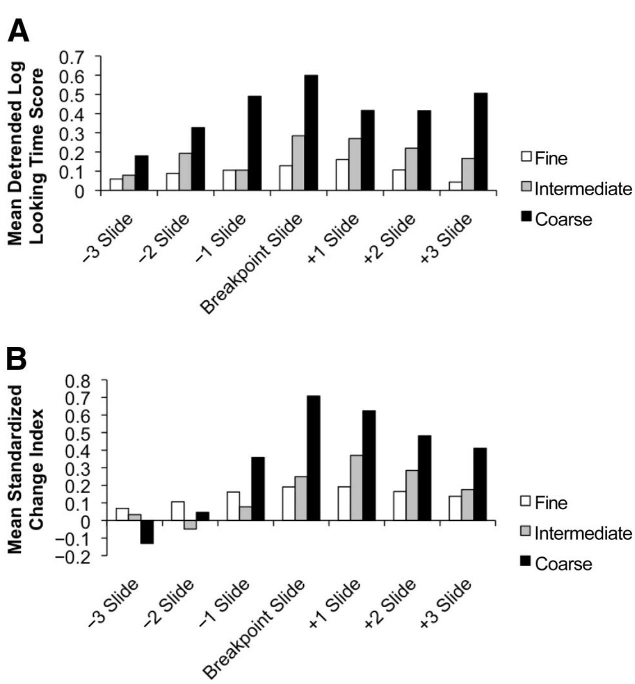

Figure 4. A: Mean detrended log looking time score in the preceding and subsequent slides to breakpoints. B: Standardized change index in the preceding and subsequent slides to breakpoints.

Although physical change does not entirely account for looking time overall, can it account for looking time specifically at fine, intermediate, or coarse breakpoints? Examination of the individual contributions of fine, intermediate, and coarse breakpoints to mean looking time indicated that at the group level, fine breakpoints predicted mean looking time (R^2 = .02, β = .13), $t(1)$ = 3.45, as did intermediate (R^2 = .03, β = .18), $t(1)$ = 4.53, and coarse (R^2 = .07, β = .26), $t(1)$ = 6.87, breakpoints. At the group level, the change value for each slide significantly predicted the likelihood that a slide was selected as a fine (R^2 = .05, β = .23), $t(1)$ = 5.96, an intermediate (R^2 = .03, β = .19), $t(1)$ = 4.80, or a coarse (R^2 = .08, β = .28), $t(1)$ = 7.31, breakpoint. Controlling for change, the likelihood that a slide was a fine breakpoint did not predict looking time, $t(1)$ = 1.61, p = .11, but the likelihood that a slide was an intermediate breakpoint did (β = .12), $t(1)$ = 3.13), and did explain significant additional variance in looking time (change R^2 = .01), $F(1, 651) = 9.79$. The likelihood that a slide was a coarse breakpoint also remained a significant predictor of looking time controlling for change (β = .18), $t(1) = 4.81$, and explained significant additional variance in looking time (change R^2 = .03), $F(1, 651) = 23.13$. These results indicate that participants looked longer at intermediate and coarse breakpoints only partly because of greater physical change, as it was measured here.

In the data for individual participants, looking time at coarse breakpoints was greater than was looking time at intermediate breakpoints. Was this solely because coarse breakpoints involved more physical change than did intermediate breakpoints? According to the regression coefficients reported above, coarse breakpoints did lead to larger increases in looking time than did intermediate breakpoints, even controlling for change. Specifically, if the number of coarse breakpoints selected for a given slide increased by 1 standard deviation, then the mean looking time for that slide increased by 0.18 of a standard deviation. Increasing the number of intermediate breakpoints selected for a given slide by 1 standard deviation only increased mean looking time by 0.12 of a standard deviation. Although these regression coefficients offer only a rough approximation of how much intermediate and coarse breakpoints influence looking time, their values suggest that participants looked longer at coarse breakpoints than at intermediate breakpoints, even controlling for change.

Together, these analyses show that although looking time tracks physical change, looking times and the physical change measure deviate in two ways. First, looking time at coarse and intermediate breakpoints is only partially explained by physical change. Second, looking time increased for the slides prior to breakpoints and prior to increases in physical change. The physical change measure captures only bottom-up quantitative information and captures that only roughly. The discrepancies between physical change and looking time may suggest limitations to our measure of physical change or that more subtle or conceptual changes at breakpoints also affect looking time.

Does looking time at breakpoints predict recall? Participants' free recall of the slideshow was coded by Gabriel Recchia by counting the number of actions reported. In some cases, an action was indicated by a phrase with a single verb and direct object, as in "she ate a banana." In other cases, multiple actions were conveyed in a single phrase, as in "she took three bites." Participants were credited for as many actions as the description conveyed. For example, "she took three bites" was coded as three actions. On average, participants recalled 23.43 actions (*SEM* = 1.62). One participant recalled 51 actions, more than 2.5 standard deviations above the mean number of actions recalled. This participant was excluded from the following analysis as an outlier. Errors in recall were not analyzed because they were so rare: Only 11 out of 863 (1.3%) total recalled actions could be considered errors.

Participants who looked longer at coarse breakpoints recalled significantly more actions later, $r(39) = .46$. Total looking time across the slideshow did not correlate with recall, $r(39) = .23$, $p = .16$, nor with looking time at coarse breakpoints, $r(39) = -.03$, $p = .87$; thus the relation between coarse breakpoint looking times and recall was not because participants with long coarse breakpoint looking times spent more time with the slideshow overall. Looking time scores at other breakpoints and within units also did not predict recall. This result indicates that selectively looking at coarse breakpoints helped participants build a memorable representation of ongoing action. But why did this occur? Did looking time at coarse breakpoints reflect encoding of coarse-level actions, or were coarse breakpoints convenient resting points for participants to consciously rehearse actions that had occurred up to that point? The latter interpretation seems unlikely because the average looking time to coarse breakpoint slides was only slightly over a second (M = 1,028 ms, *SEM* = 115.2 ms), compared with the average time per slide of slightly less than a second (M = 795 ms, *SEM* = 68.7 ms). One second seems insufficient to actively rehearse a long list of actions. Rather, it seems that selectively looking at coarse breakpoints augmented high-level understanding and integration of the action sequence.

Does looking time at breakpoints predict hierarchical segmentation? If looking selectively at coarse breakpoints facilitates understanding and integration, then it may also facilitate perceiving hierarchical structure during the segmentation task. Hierarchical segmentation was assessed with both quantitative and qualitative indices. The quantitative index, *alignment*, is based on the premise that when coarse and fine units of action are hierarchically related, their boundaries coincide in time (Zacks, Tversky, & Iyer, 2001). Boundaries of coarse units, like *putting on the top sheet*, should coincide with the boundaries of relevant subunits, like *tucking in the (last) corner*. Here, we adopted the continuous method for assessing alignment developed by Zacks, Tversky, and Iyer (2001) and, following Hard et al. (2006), calculated a set of alignment scores for each participant. A detailed description of the calculations appears in Appendix B. Separate alignment scores were calculated between intermediate and coarse units (intermediate-coarse alignment) and between fine and intermediate units (fine–intermediate alignment). Intermediate–coarse alignment (M = .63, *SEM* = .04) was significantly greater than fine–intermediate alignment (M = .53, *SEM* = .04), paired $t(39)$ = 2.10, d = 0.40. This difference is probably due to the fact that fine breakpoints occur more frequently than do coarse breakpoints, so that segmenting activities into fine units is more demanding and error-prone than segmenting activities into coarse units, decreasing alignment.

Table 1 shows the correlations between looking time at breakpoints and subsequent alignment scores. Participants who looked longer at fine breakpoints later had higher fine–intermediate alignment. Participants who looked longer at within-unit slides, in contrast, had lower fine–intermediate alignment. Together, these findings suggest that participants who focus their attention selectively on breakpoints later segment the action sequence in a more organized fashion.

Surprisingly, looking time at intermediate and coarse breakpoints predicted neither fine–intermediate alignment nor intermediate–coarse alignment. Why does looking time at fine breakpoints, but not intermediate and coarse breakpoints, predict alignment? This pattern of results may be interpretable if we consider that high alignment scores likely reflect high reliability in breakpoint perception, that is, a tendency to mark fine breakpoints at reliably the same temporal location as corresponding breakpoints at coarser levels. Participants who pay close attention to activity at a fine level may be more tuned to specific changes in the flow of action and use those changes consistently as indicators of breakpoints at any level.

Table 1
Correlation Coefficients Between Looking Time and Measures of Alignment and Enclosure

	Fine–intermediate		Intermediate–coarse		Average	
LT	Alignment	Enclosure	Alignment	Enclosure	Alignment	Enclosure
Coarse LT	.16	.27†	.14	.35*	.18	.45*
Intermediate LT	−.15	.05	−.20	.24	−.21	.12
Fine LT	.46*	.18	.05	.01	.29†	.18
Within-unit LT	−.44*	.03	−.09	−.13	−.30†	−.06

Note. Looking time (LT) scores were measured at breakpoints (coarse, intermediate, and fine) in the slideshow task, whereas alignment and enclosure scores were measured during the segmentation task.
† $p < .10$. * $p < .05$.

The qualitative index of hierarchical segmentation, *enclosure*, is based on the premise that when coarse and fine units of action are hierarchically related, coarse units *contain* fine ones. If so, boundaries of coarse units should be marked after or outside of, that is, enclosing, boundaries of fine units. A detailed description of enclosure calculation appears in Appendix B. Enclosure was calculated for each participant between intermediate and coarse breakpoints (intermediate–coarse enclosure) and between fine and intermediate breakpoints (fine–intermediate enclosure). Intermediate–coarse enclosure (M = .65, SEM = .03) was significantly higher than was fine–intermediate enclosure (M = .55, SEM = .03), paired $t(39) = 2.20$, $d = 0.51$. This finding may again be explained by the heightened demands and noisier data for fine-level segmentation.

Table 1 shows the correlations between looking time scores at breakpoints and later enclosure scores. In contrast to alignment scores, enclosure scores correlated with looking time at coarse breakpoints but not at finer breakpoints. An interpretation of this finding is that looking time at coarse breakpoints is related to the integration of component actions into larger wholes, leading observers to later perceive coarser units of action as containing the finer units.

How was participants' memory of actions from the slideshow related to their later segmentation patterns? The number of actions that participants recalled was significantly correlated with average enclosure score, $r(37) = .36$, and marginally correlated with average alignment scores, $r(37) = .29$, $p = .08$. These findings reinforce that better memory for an activity is associated with a more organized perception of that activity in an explicit segmentation task.

Discussion

Organized human activity unfolds continuously in time, yet is thought of as segments that are organized hierarchically. But is online perception of human activity segmented and hierarchical? Previous work suggested that it is (Zacks, Braver, et al., 2001; Zacks, Tversky, & Iyer, 2001), and the present results add substantially to that evidence. In the present study, participants looked at slides depicting a human activity at their own pace, yielding a measure of looking time or attention. Later, they recalled the actions they had viewed and then segmented films of the same activity into breakpoints at several levels. Slide-to-slide change was also computed. Earlier research suggests that breakpoints involve the convergence of perceptual and conceptual changes in action. Thus, we predicted that breakpoints contain more information to process and so should be marked at moments of greater change and should be looked at longer. Support for these ideas was striking: Change in action and looking time rose and fell together, with breakpoints marked at their peaks. What is more, both change and attention were modulated by level within the action hierarchy: Coarse level breakpoints involved more physical change and attracted more attention than did intermediate breakpoints, and intermediate, more than fine level breakpoints.

The close correspondence between perceived hierarchical structure and low-level physical change also suggests that a measure of relative physical change can serve as an excellent cue to the hierarchical structure of the activities. This finding raises the possibility of automatically processing films of human activity for their perceived hierarchical organization. The deeper significance is that automatic processing of activities based on relative physical change would reveal not only the perceptual structure of activities but also at least the rough outlines of the conceptual structure.

Although looking time and physical change were closely coupled, especially at breakpoints, the time courses of looking time and change diverged in a suggestive way. The second-by-second analyses showed that prior to breakpoints, looking time began to increase a few seconds before increases in physical change. The fact that increases in looking time occur prior to increases in physical change suggests that subtle changes not captured by our change measure may be cuing an imminent goal completion and capturing viewers' attention in advance of the breakpoint. Actors typically prepare for the next action as they finish the current one, shifting their gaze to the next object to be acted on (Mennie, Hayhoe, & Sullivan, 2007). Directing the gaze to an object is an early cue of intention to act on that object (Baird & Baldwin, 2001; Pierno et al., 2006). Experienced observers of human action are likely to notice those subtle yet reliable cues to an upcoming transition and may increase attention in anticipation of the next set of actions.

The correspondences between looking times and later measures of hierarchical segmentation suggest that looking times can serve as an implicit measure of segmentation and of hierarchical perception of human action. Looking time was compared with two indices of hierarchical segmentation: alignment and enclosure. Each of these measures correlated with looking time at breakpoints, but alignment was more strongly correlated with looking time at fine breakpoints, and enclosure was more strongly correlated with looking time at coarse breakpoints. Alignment measures how breakpoints of coarser and finer units fall closely in time, irrespective of whether finer breakpoints are perceived before or after corresponding coarser breakpoints. Degree of alignment may reflect degree of consistency of the features participants consciously use to mark breakpoints. High consistency may be associated with greater attention to fine levels of detail and thus longer looking at fine breakpoints. In contrast, enclosure measures a particular temporal ordering of closely aligned breakpoints: Larger units of action cannot be complete until all the component actions are complete, such that finer unit boundaries fit within intermediate unit boundaries that fit within coarse unit boundaries. We propose that such temporal relations reflect integration of related actions into chunks. Chunking of actions may be associated with greater attention to the overarching goals of the activity and thus longer looking at coarse breakpoints. An implication of this analysis is that degree of enclosure may reflect degree of hierarchical organization better than degree of alignment. Further research is needed to explore that possibility.

Importantly, selective looking at coarse breakpoints and not total looking time was associated with greater recall. These findings suggest that techniques that increase hierarchical perception of ongoing activity should increase learning and memory. As such, the findings have important implications for action learning, which typically occurs through observation and imitation.

Experiment 2: Breakpoints Are Looked at Longer out of Context

So far, the evidence confirms that breakpoints are significant because they are informative within the context of the flow of action—they correspond to peaks of physical change that demand attention. Breakpoints can thus be likened to mountain peaks of new information separated by valleys. But are breakpoints also informative out of context? As suggested earlier, breakpoints are also bridges from one action to another. Because transitions between actions can involve considerable overlap of those actions (Mennie, Hayhoe, & Sullivan, 2007), breakpoints can simultaneously capture the completion of one action and the initiation of the next. In such moments of transition, the actor's body and objects being acted on might take on configurations that are informative even when presented in isolation. If so, then breakpoints should capture attention even out of context.

Our first goal in this next experiment was to examine the role of temporal context in looking time by scrambling the order of the action sequences for half the participants. A second goal was to test whether viewers attend selectively to breakpoints, even when the breakpoints are defined by other people and occur as frequently as within-unit moments. New slideshows were constructed from the previous ones based on the data collected in Experiment 1. Half the slides were ones judged by Experiment 1 participants as breakpoint moments at each of the three segmentation levels, and half were ones judged as within-unit moments. New participants viewed these slideshows, with half seeing the slides in chronological order and the other half seeing the slides in random order. If breakpoints are especially informative only in context, then breakpoints should attract longer looking times under chronological presentation but not under scrambled presentation. On the other hand, if breakpoint moments are inherently distinctive and informative, they should attract longer looking times even in the absence of coherent temporal sequencing.

Method

Stimuli. The same four activities were used in this study as in Experiment 1. Sixty slides from each activity were selected based on the segmentation data from Experiment 1, so that exactly half ($N = 30$) corresponded to breakpoints and half corresponded to within-units ($N = 30$). Slides were selected based on the average likelihood that each slide had been selected by participants in Experiment 1 as a breakpoint. The 10 slides with the highest likelihood of being selected as a coarse, intermediate, or fine breakpoint in Experiment 1 were used as the coarse, intermediate, and fine breakpoint slides (respectively) in this study. The categories were mutually exclusive; that is, a slide could not be selected for the new slideshow as an intermediate breakpoint if it had already been selected as a coarse breakpoint. Slides could not be fine breakpoints if already selected as intermediate or coarse breakpoints. All slides selected for the present study had average likelihoods at least 1 standard deviation above the mean. Within-unit slides were selected to be the 30 slides with the lowest likelihood of being selected as any breakpoint (the sum of coarse, intermediate, and fine). Before being used in the experiment, the slides selected for the makeup slideshow were checked to see whether they were free of digital "scratches," so that they could be included in analyses of physical change. Only six slides out of 60 slides contained small scratches (three intermediate breakpoint slides and three within-unit slides), so we decided to include them and the makeup slideshow in the experiment and in further analyses. We did confirm, however, that including or excluding these six slides from the analyses did not change the pattern or reliability of the results.

Participants and design. Forty Stanford undergraduates participated in the study in exchange for pay or course credit. One participant did not follow instructions and was replaced. The experimental design was a $4 \times 4 \times 2$ mixed factorial design. Presentation order (scrambled, chronological) was varied between subjects. Slide type (fine breakpoint, intermediate breakpoint, coarse breakpoint, and within unit) was varied within subjects. Because the slideshows for this study were shorter than the ones used in Experiment 1, participants were asked to watch all four, rather than just one. Thus, activity (cleaning, breakfast, makeup, and TV cart) was also varied within subjects.

Procedure. Participants were seated at a computer and told that they would view several slideshows. Participants were told that immediately after a slideshow ended, the computer would prompt them to write down as many actions as they could remember before moving on to the next slideshow. They were asked not to write down actions while watching the slideshow and not to change answers after moving to the next slideshow. Participants in the scrambled condition were told that the slides would be in a scrambled sequence and that they should try their best to make sense of them. As before, slideshow viewing was self-paced, and participants pressed the spacebar to advance to the next slide.

Participants then viewed all four slideshows (cleaning, breakfast, makeup, or TV cart) in a randomized order. In the chronological condition, slides within each slideshow were presented chronologically. In the scrambled condition, the order of slides within each slideshow was randomized for each individual participant. Immediately after each slideshow, the computer prompted participants to recall as many actions as they could remember from that slideshow before moving on to the next one. The instructions for recall were the same as in Experiment 1.

Results

How does presentation order affect looking time? As for the first experiment, the looking time data were first transformed with a $\log_{10}$ function and detrended with power functions for each individual participant. Looking times were again well described by a power function, accounting for an average of 29% of the variance in looking time in the chronological condition (SEM = .03) and 32% of the variance in the scrambled condition (SEM = .03). The variance accounted for by the power function did not differ significantly between the two conditions, $t(38) < 1$. From the detrended data, average looking times were then calculated for each participant. Separate average looking times were calculated for fine breakpoints, intermediate breakpoints, coarse breakpoints, and within units and were averaged across the four slideshows. Unlike in Experiment 1, looking time scores were not calculated by dividing these average looking times by their standard deviation because there were equal numbers of breakpoint and within-unit slides, and equal numbers of coarse, intermediate, and fine breakpoint slides.

Average looking times were first submitted to a mixed factorial ANOVA with slide type (breakpoint, within unit) as a within-subjects factor and presentation order (chronological, scrambled) as a between-subjects factor. First, participants looked longer at breakpoint slides (M = 0.05, SEM = 0.005) than at within-unit slides (M = −0.03, SEM = 0.004), $F(1, 38) = 121.15$, MSE = .001, $\eta_p^2 = .76$. Second, there was a small main effect of presentation order, $F(1, 38) = 7.37$, $MSE = 2.59 \times 10^{-5}$, $\eta_p^2 = .16$, such that participants had higher looking times in the chronological condition (M = 0.0093, SEM = 0.00084) than in the scrambled condition (M = 0.0062, SEM = 0.00077). This effect requires clarification: Recall that the analyzed looking times for each slide are transformed and detrended versions of the original looking time for that slide. Although participants in the chronological condition had higher transformed and detrended looking times on average, analysis of the raw looking times indicated that participants in the scrambled condition looked at each slide on average nearly twice as long (M = 2,193 ms, SEM = 300 ms) as participants in the chronological condition (M = 1,106 ms, SEM = 97 ms), $t(38) = 3.45$, $d = 1.12$.

If the raw looking times were, on average, longer in the scrambled condition than in the chronological condition, then why did the detrended looking times, on average, appear to be longer in the chronological condition? Recall that the detrending process removes variance that is attributable to the overall decrease in looking time as participants become accustomed to the slideshow. What is left after the detrending is variance attributable to other factors, including whether the slide was a breakpoint or not. In the chronological slideshow, the high detrended looking times specifically at breakpoints seemed to drive up the average detrended looking times in that condition. This was observed in an interaction between presentation order and slide type, $F(1, 38) = 36.52$, $\eta_p^2 = .49$. As evident in Figure 5, the difference in looking time between breakpoint slides and within-unit slides was greater in the chronological condition, $F(1, 19) = 102.89$, $MSE = 1.34 \times 10^{-3}$, $\eta_p^2 = .84$, than in the scrambled condition, $F(1, 19) = 20.97$, $MSE = 5.59 \times 10^{-4}$, $\eta_p^2 = .53$. Analysis of the looking time data for individual participants indicated that all 20 participants in the chronological condition had longer looking times at breakpoints than within units. Sixteen of the 20 participants in the scrambled condition showed this difference. The difference in the frequencies of participants showing this effect significantly differed between conditions, $\chi^2(1, N = 40) = 4.44$. We confirmed that the effect of slide type and other subsequently reported effects on looking time held, whether analyzing the transformed and detrended looking times or raw looking times.

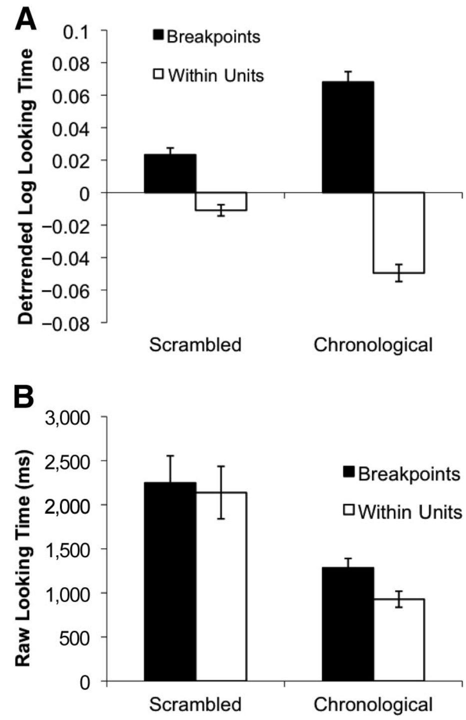

Figure 5. Mean looking time for slides categorized as breakpoints and as within units in Experiment 2, as a function of presentation order. Error bars reflect the standard error of the mean. A: The top graph shows the transformed and detrended versions of the looking times that are reported in the analyses. B: The bottom graph shows the raw looking times in milliseconds.

Mean looking times for breakpoint slides were next submitted to a mixed factorial ANOVA with segmentation level (coarse, intermediate, fine) as a within-subjects factor and presentation order (chronological, scrambled) as a between-subjects factor. In Experiment 1, looking time varied as a function of segmentation level. This was true in the present study also, $F(2, 76) = 37.98$, $MSE = 1.83 \times 10^{-3}$, $\eta_p^2 = .50$. Looking time at breakpoints showed a linear trend, such that coarse breakpoints were looked at longest (M = 0.09, SEM = 0.01), followed by intermediate (M = 0.04, SEM = 0.01) and then fine breakpoints (M = 0.01, SEM = 0.01), $F(1, 38) = 51.99$, $\eta_p^2 = .58$.

The effect of segmentation level interacted with presentation order, $F(2, 76) = 17.69$, $\eta_p^2 = .32$ (see Figure 6). Separate

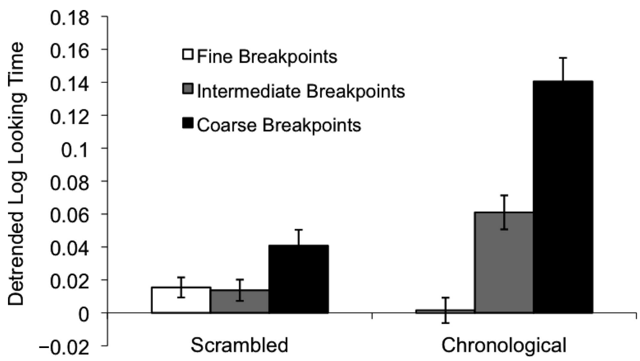

Figure 6. Mean looking time at breakpoints in Experiment 2 as a function of segmentation level and presentation order. Error bars reflect the standard error of the mean.

analysis of the two presentation order conditions revealed a strong effect of segmentation level in the chronological condition, $F(2, 38) = 39.64$, $MSE = 2.45 \times 10^{-3}$, $\eta_p^2 = .68$. This effect was linear; looking times at coarse breakpoints ($M = 0.14$, $SEM = 0.01$) were highest, followed by intermediate ($M = 0.06$, $SEM = 0.01$) and fine breakpoints ($M = 0.002$, $SEM = 0.01$), $F(1, 19) = 51.93$, $MSE = 3.72 \times 10^{-3}$, $\eta_p^2 = .73$. There was a weaker effect of segmentation level in the scrambled condition, $F(2, 38) = 3.81$, $MSE = 1.21 \times 10^{-3}$, $\eta_p^2 = .17$. This effect was also characterized by a linear trend, albeit a weaker one, $F(1, 19) = 4.35$, $MSE = 1.47 \times 10^{-3}$, $p = .05$, $\eta_p^2 = .19$: As Figure 6 shows, looking times were highest at coarse breakpoints ($M = 0.04$, $SEM = 0.01$), but intermediate ($M = 0.01$, $SEM = 0.01$) and fine breakpoints ($M = 0.02$, $SEM = 0.01$) had similar looking times. Analysis of the trends for individual participants indicated that 15 of the 20 participants in the chronological condition showed the predicted linear pattern of looking time (coarse > intermediate > fine), but only four of the 20 participants in the scrambled condition showed the same pattern. The difference in the frequencies of participants showing this effect significantly differed between conditions, $\chi^2(1, N = 40) = 12.13$.

How does presentation order affect change? Experiment 1 showed that breakpoints corresponded to relatively large amounts of physical change and that these changes increased with level of organization. Was this true in Experiment 2 as well, in which each participant saw only a sample of slides from Experiment 1 and in which another group of participants defined the breakpoints? Are breakpoint slides more physically different from the previous slide than ordinary slides are, even in scrambled order?

A consideration before analyzing physical change is that the slides in this experiment were only a sampling of still frames from the original slideshow used in Experiment 1. Whereas the temporal distance between successive slides in Experiment 1 was necessarily 1 s, the temporal distance between successive slides in this slideshow could have been a second or much longer. It is possible that in this new slideshow, the breakpoint slides that we selected might, on average, have been further away from the preceding slide, in terms of true temporal distance, than were within-unit slides. If so, then breakpoint slides in the scrambled condition might be physically more different from their preceding slide simply because they were further away in time. We checked this by calculating the amount of true time (in s) separating each slide used in the reduced Experiment 2 slideshow from its preceding slide, based on the complete version of the slideshow in Experiment 1. There were no differences between breakpoint slides ($M = 2.7$, $SEM = 0.21$) and within-unit slides ($M = 2.5$, $SEM = 0.19$), $t(234) < 1$, and no differences among breakpoint slides at different levels of abstraction, $F(3, 235) < 1$.

In Experiment 1, changes at breakpoint and within-unit slides were calculated individually for each participant because each participant identified subtly different breakpoints within the sequence. In the present study, each participant in the chronological condition viewed exactly the same sequence of within-unit and breakpoint slides, whereas each participant in the scrambled condition viewed a different random sequence. Thus, a different analysis of change was required. Each slide was treated as a "subject," and a chronological change value for each slide was calculated with the same technique used in Experiment 1. For each slide, a change value was then calculated for each participant in the scrambled condition, based on the specific, random temporal ordering of slides that the participant saw. The slides and their corresponding change values were then reordered in the correct chronological sequence and averaged across all of the participants in the scrambled condition, yielding an average scrambled change value for that slide.

Change values were standardized for each activity for ease of interpretation and because, as in Experiment 1, the change values varied across activity, both for the chronologically presented slideshows, $F(3, 231) = 13.00$, $MSE = 2.90 \times 10^{11}$, $\eta_p^2 = .14$ and the scrambled slideshows, $F(3, 231) = 106.74$, $MSE = 1.48 \times 10^{11}$, $\eta_p^2 = .58$. As before, activity did not interact with any of the effects presented here. The standardized change values were submitted to a mixed factorial ANOVA, with presentation order (chronological, scrambled) as a within-subjects factor and slide type (breakpoint, within unit) as a between-subjects factor. Standardizing the change values equates the mean amount of change within the scrambled and chronological conditions. Notably, the mean amount of change in the scrambled and chronological conditions was not equal. Overall, there was more change in the scrambled condition ($M = 2,333,213$; $SEM = 38,475$) than in the chronological condition ($M = 1,569,775$; $SEM = 37,779$), paired $t(234) = 23.58$, $d = 1.31$. This was expected given that for each slide in the chronological sequence, the preceding slide corresponded to a point in the original video about 2.6 s away in time, on average. For each slide in the scrambled sequence, the preceding slide could correspond to points in time hundreds of seconds away, when the actor's body position was likely dramatically different. This is also consistent with the finding of increased looking time overall for the scrambled slides.

Analysis of the standardized change values indicated that across the two presentation orders, breakpoint slides corresponded to greater amounts of change ($M = 0.41$, $SEM = 0.09$) than within-unit slides ($M = -0.41$, $SEM = 0.07$), $F(1, 233) = 42.85$, $MSE = 1.33$, $\eta_p^2 = .16$. This effect interacted with presentation order, $F(1, 233) = 8.51$, $MSE = 0.38$, $\eta_p^2 = .04$ (see Figure 7). Separate analysis of the chronological and scrambled change values indicated a much larger effect of slide type in the chronological condition (for breakpoints: $M = 0.44$, $SEM = 0.09$; for within units: $M = -0.43$, $SEM = 0.08$), $F(1, 233) = 54.39$, $MSE = 0.80$,

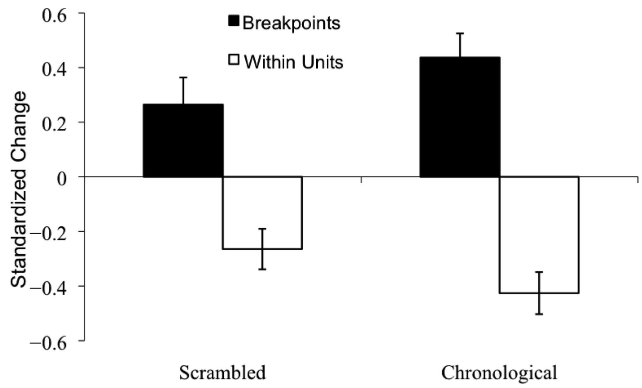

Figure 7. Mean standardized change values in Experiment 2 as a function of slide type and presentation order. Error bars reflect the standard error of the mean.

η_p^2 = .19, than in the scrambled condition (for breakpoints: *M* = 0.26, *SEM* = 0.10; for within units: *M* = −0.26, *SEM* = 0.07), *F*(1, 238) = 18.35, *MSE* = 0.91, η_p^2 = .07. Thus, even when presented in scrambled order, breakpoint slides contained relatively large changes relative to other parts of the behavior stream. This suggests that breakpoint slides are distinctive—they look more different from a randomly selected slide than within-unit slides do, even when the numbers of breakpoint and within-unit slides are equated.

Change values for breakpoint slides were next submitted to a mixed factorial ANOVA, with presentation order (chronological, scrambled) as a within-subjects factor and segmentation level (coarse, intermediate, fine) as a between-subjects factor. There was a main effect of segmentation level, *F*(2, 113) = 4.39, *MSE* = 1.59, η_p^2 = .07. Coarse breakpoints had the highest change (*M* = 0.83, *SEM* = 0.17), followed by intermediate breakpoints (*M* = 0.34, *SEM* = 0.14) and then fine breakpoints (*M* = 0.06, *SEM* = 0.14).

There was also a significant interaction between segmentation level and presentation order, *F*(2, 113) = 4.17, *MSE* = 0.37, η_p^2 = .06 (see Figure 8). Separate analysis of the chronological and scrambled conditions indicated a reliable effect of segmentation level in the chronological condition, *F*(2, 113) = 6.43, *MSE* = 0.83, η_p^2 = .10, but only a marginally reliable effect in the scrambled condition, *F*(2, 117) = 2.84, *MSE* = 1.14, *p* = .06, η_p^2 = .05. The effect in the chronological condition was well characterized by a linear trend: Coarse breakpoints had the greatest amount of change (*M* = 0.80, *SEM* = 0.16), followed by intermediate breakpoints (*M* = 0.49, *SEM* = 0.14) and then fine breakpoints (*M* = 0.06, *SEM* = 0.15), *F*(2, 113) = 6.43, *MSE* = 0.83, η_p^2 = .10. For the scrambled condition, there was only a marginally reliable linear trend, *F*(2, 117) = 2.96, *MSE* = 1.15, *p* = .06, η_p^2 = .05: Visual inspection of Figure 8 shows how the pattern of looking time differed in the scrambled condition, compared with the chronological condition: Coarse breakpoints had the highest degrees of change (*M* = 0.59, *SEM* = 0.20), but intermediate breakpoints did not have higher change (*M* = 0.04, *SEM* = 0.16) than fine breakpoints (*M* = 0.16, *SEM* = 0.14).

The fact that breakpoints corresponded to both heightened attention and heightened change, even in scrambled sequences of slides, is surprising and suggests that breakpoint moments are inherently distinctive. If breakpoint slides really are informative, even out of context, are they more different from every other slide in the slideshow? Once again using our metric for determining the similarity of two slides, each slide in the present experiment was compared with every other slide within the same slideshow. Because there were 60 slides in each of the four slideshows, these comparisons yielded a set of 59 change values for each slide. These 59 change values were averaged, yielding a single number for each slide that represented the global difference between that slide and all other slides in the slideshow. The global difference values were first standardized for each activity for ease of interpretation and because, as for the change values reported in Experiment 1 and previously for this experiment, the average global difference values varied across the four activities, *F*(3, 236) = 124.61, *MSE* = 1.39, η_p^2 = .61. Activity did not interact with any of the subsequent effects reported here. The standardized global difference values were next submitted to an independent-samples *t* test with slide type (breakpoint, within-unit) as a between-subjects factor. Breakpoint slides were more different (*M* = 0.28, *SEM* = 0.10), compared with all other slides than were within-unit slides (*M* = −0.28, *SEM* = 0.07), *t*(238) = 4.61, *d* = 0.59. The standardized global difference values were next submitted to a one-way ANOVA with breakpoint type (fine, intermediate, coarse) as a factor. This analysis indicated that the global differences varied with segmentation level, *F*(2, 117) = 3.18, *MSE* = 1.23, η_p^2 = .05. On the basis of the pattern of data we observed for physical change in the scrambled condition, we predicted that coarse breakpoints would likely drive any effect of segmentation level on the global difference measure. To test this prediction, two planned contrasts were performed. The first indicated that coarse breakpoints had significantly higher global differences (*M* = 0.64, *SEM* = 0.20) than intermediate and fine breakpoints, *t*(117) = 2.51. The second showed that intermediate (*M* = 0.07, *SEM* = 0.17) and fine breakpoints (*M* = 0.14, *SEM* = 0.15) did not significantly differ from one another, *t*(117) < 1.

Explaining looking time: A regression analysis. Experiment 1 showed that physical changes at breakpoints play a significant role in why observers look longer at breakpoints. But participants looked at breakpoints, especially coarse breakpoints, for reasons besides change. Was this true in Experiment 2 as well?

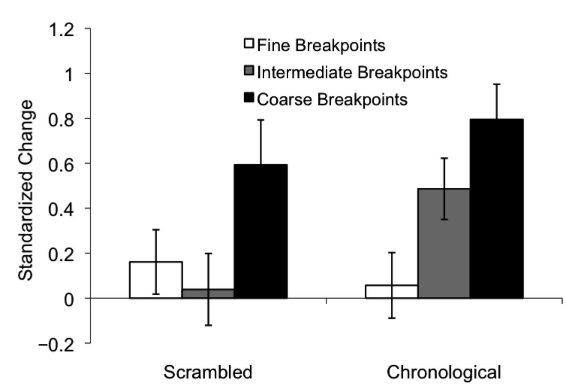

Figure 8. Mean standardized change values at breakpoints in Experiment 2 as a function of segmentation level and presentation order. Error bars reflect the standard error of the mean.

To find out, the mean looking time in ms was determined for each slide by pooling and averaging the looking times of the 20 participants in the chronological condition who viewed that slide. The mean looking times for each slideshow were then transformed with a $\log_{10}$ function to reduce positive skewness and were detrended by fitting a power function and extracting the residuals. These values were then standardized for each slideshow and used as the dependent measure in a regression analysis.

Whether a slide was a breakpoint or not (coded as binary: 1 or 0) significantly predicted mean looking time in the chronological condition ($R^2 = .18$, $\beta = .43$), $t(1) = 7.34$, as did the change value for each slide ($R^2 = .36$, $\beta = .60$), $t(1) = 11.44$, and the global difference value for each slide ($R^2 = .15$, $\beta = 0.39$), $t(1) = 6.47$. Controlling for change, the global difference value no longer predicted looking time significantly, ($\beta = -.05$), $t(1) < 1$. Controlling for change, being a breakpoint did significantly predict looking time ($\beta = .21$), $t(1) = 3.77$, and explained significant additional variance above and beyond change (change $R^2 = .04$), $F(1, 232) = 14.20$. This replicates findings from Experiment 1 that physical change accounts for some, but not all, of the relation between breakpoints and looking time when slides are viewed chronologically.

Examination of the individual contributions of fine, intermediate, and coarse breakpoints to looking time indicated that at the group level, being a fine breakpoint (coded as binary, 1 or 0) was not a significant predictor of mean looking time ($R^2 = 0$, $\beta = -.004$), $t(1) < 1$, $p = .95$), but being an intermediate ($R^2 = .02$, $\beta = .15$), $t(1) = 2.39$, or coarse ($R^2 = .18$, $\beta = .43$), $t(1) = 7.28$, breakpoint was. Controlling for change, being an intermediate breakpoint did not significantly predict looking time ($\beta = .02$), $t(1) < 1$, but being a coarse breakpoint did ($\beta = .24$), $t(1) = 4.46$. Being a coarse breakpoint predicted significant additional variance above and beyond the contribution of change (change $R^2 = .05$), $F(1, 232) = 19.88$. These results are consistent with those of Experiment 1.

Did these results hold in the scrambled condition? Recall that the 20 participants in the scrambled condition each saw the slides in a different random sequence. Thus, before a regression analysis could be performed, each participant's looking times were reordered so that the slides would be in chronological sequence. The looking times for each slide were then averaged across the 20 participants in the scrambled condition and then $\log_{10}$ transformed to reduce skewness. Although each individual participant's looking times were well characterized by a power function for the specific random sequence of slides that he or she saw, once the looking times were reordered and averaged across the 20 participants, they were no longer characterized by a power function, making the detrending process unnecessary.[5] The looking time values were then standardized for each activity and used as a dependent measure in a regression analysis.

Whether a slide was a breakpoint did not predict mean looking time in the scrambled condition ($R^2 = .01$, $\beta = .09$), $t(1) = 1.31$, $p = .19$, but change values did ($R^2 = .03$, $\beta = .17$), $t(1) = 2.70$, as did the global difference values ($R^2 = .03$, $\beta = .18$), $t(1) = 2.80$. Controlling for change, the global difference values did not predict looking time, however, ($\beta = 0.14$), $t(1) < 1$. Thus, how different the slides looked relative to the immediately previous slide was the primary predictor of looking time in the scrambled condition. Notably, the relation between change and looking time in the scrambled condition was not as strong as in the chronological condition ($R^2 = .03$ vs. $R^2 = .36$). This may be because in the chronological condition, physical change was more meaningful; as physical change in the slides increased, so did the change in the actor's actions and goals. This likely made it easier for observers to detect physical changes and exploit them in processing the activity. Additionally, change was very high between all slides in the scrambled condition, so although breakpoints did correspond to heightened amounts of change, these changes were probably harder to detect.

Examination of the individual contributions of fine, intermediate, and coarse breakpoints to mean looking time indicated that at the group level, being a fine breakpoint was not a significant predictor of mean looking time ($R^2 = .003$, $\beta = .05$), $t(1) < 1$, $p = .40$, nor was being an intermediate breakpoint ($R^2 = .002$, $\beta = -0.05$, $t(1) < 1$. Being a coarse breakpoint was a marginally reliable predictor ($R^2 = .01$, $\beta = 0.11$), $t(1) = 1.66$, $p = .10$. None of these were significant predictors controlling for change.

This pattern of results suggests that the looking time effects observed in the scrambled condition were attributable to different processes than in the chronological condition. In the chronological condition, looking time was attributable largely to physical changes in the behavior stream, but there was additional unexplained variance based on whether participants were looking at a breakpoint, especially a coarse breakpoint. In the scrambled condition, looking time had a very simple explanation: The only significant predictor of looking time was the physical changes between slides. No additional variance was attributable to whether participants were looking at a breakpoint.

Does looking time predict recall? In the first study, looking time at coarse breakpoints was correlated with later recall. Were looking time and recall related in the present study, and did their relation depend on presentation order? In this study, Gabriel Recchia, blind to condition (chronological, scrambled), determined the number of actions recalled for each slideshow the same way as in Experiment 1 and then averaged across the four slideshows. In contrast to Experiment 1, errors in recall were frequent enough to merit analysis, which is unsurprising given that participants had to keep track of four different activities and saw only incomplete slides from each. Overall, participants recalled an average 13.79 correct actions for each slideshow (*SEM* = 0.68) and made 0.36 recall errors (*SEM* = 0.05). As in memory for scripts (e.g., Bower et al., 1979), most of the recall errors (about 70%) were intrusions consistent with the theme of the activity, like reporting that the woman in the cleaning slideshow had opened the lid of her jewelry box and looked inside, when she had only moved it to a shelf. Intrusions were likely common because participants had to rely on inferences using top-down knowledge to make sense of the incomplete slideshows they were observing. The remaining errors (30%) appeared to be intrusions from the wrong slideshow, like reporting that the woman in the cleaning slideshow had answered her cell phone, when it was the woman in the breakfast slideshow who had done that. Participants in the chronological condition recalled more

[5] An alternative approach is to take the detrended looking times for participants in the scrambled condition, reorder them chronologically, and then average them across participants to get an average detrended looking time for each slide. This approach does not change the pattern of results.

correct actions (M = 15.13, SEM = 1.09) than did participants in the scrambled condition (M = 12.45, SEM = 0.73), $t(38)$ = 2.05, d = 0.67. Participants in the scrambled condition made more recall errors (M = 0.53, SEM = 0.08) than did participants in the chronological condition (M = 0.18, SEM = 0.05), $t(38)$ = 3.97, d = 1.29.

Overall, mean looking time at breakpoints positively predicted the number of actions recalled, $r(38)$ = .35 (see Table 2). A separate look at each segmentation level revealed that only looking time at coarse breakpoints predicted recall, $r(38)$ = .37. Looking time at fine and intermediate breakpoints did not predict recall, $rs(38)$ = −.07 and .18, respectively. Looking time within units were marginally negatively correlated with recall, $r(38)$ = −.30, p = .06. Errors in recall were correlated with looking time in several ways. First, looking time within units and at fine breakpoints predicted more recall errors, $rs(38)$ = .49 and .33, respectively. Looking time at intermediate and coarse breakpoints predicted fewer recall errors, $rs(38)$ = −.49 and −.53, respectively.

On the surface, these correlations replicate findings from Experiment 1, but they must be interpreted carefully. Bear in mind that participants in the chronological condition not only recalled more actions and made fewer recall errors, they had longer looking times at coarse and intermediate breakpoints and lower looking times at fine breakpoints and within units than did participants in the scrambled condition. These correlations might be artifacts of differences between the chronological and scrambled conditions rather than a reflection of a true relation between looking time and recall. To be more confident that the correlations are not artifactual, we would need to see significant correlations when controlling for presentation order or when analyzing the two presentation orders separately.

Controlling for presentation order, all the correlations just reported followed the same trends, but not reliably (see Table 2). When each condition was analyzed separately, the correlations also followed the same trends, but not reliably. These correlations might prove to be reliable with a larger sample size, thus providing a stronger replication of findings from Experiment 1, which indicated that processing time at breakpoints is meaningful to generating a coherent representation of observed behavior.

Table 2
Correlation Coefficients Between Looking Time and Recall

LT	Correct actions recalled	Recall errors
Coarse LT	.37* (.23)	−.53* (−.26)
Chronological, scrambled	**.17**, **.35**	**−.37**†, **−.18**
Intermediate LT	.18	−.49* (−.28†)
Chronological, scrambled	**−.06**, **.19**	**−.41**†, **−.18**
Fine LT	−.07	.33* (.26)
Chronological, scrambled	**−.001**, **−.01**	**.21**, **.32**
Within-unit LT	−.30†	.49* (.17)
Chronological, scrambled	**−.08**, **−.18**	**.35**, **.02**

Note. Correlation coefficients between recall and looking time (LT) were measured in the slideshow task in Experiment 2. Numbers in parentheses show correlations controlling for presentation order: These are reported in instances in which the overall correlations between recall and looking time were significant. Bold numbers reflect separate correlation coefficients for the chronological and scrambled conditions, respectively.
† $p < .10$. * $p < .05$.

Discussion

Breakpoints are privileged, even when viewed as part of a randomly ordered sequence of slides from an organized activity. Breakpoints were looked at longer than ordinary moments, though less so for scrambled than for chronological sequences. In the scrambled sequence, looking times were longer primarily at coarse breakpoints, with little distinction between fine and intermediate breakpoints. By contrast, for chronological sequences, looking time was greatest for coarse breakpoints, then intermediate, and then fine breakpoints. Critically, the breakpoints were not defined by each participant individually but were instead defined by the participants from Experiment 1. That viewers look longer at breakpoints that are not self-selected confirms that there is consistency in how viewers perceive and understand observed activities, at least activities that are likely familiar. The change measure mirrored the looking time measure in both the scrambled and chronological conditions. Thus, even in random sequences, breakpoint slides, especially coarse breakpoint slides, were more physically distinctive, locally, from preceding slides than were within-unit slides. Going further, breakpoint slides were more physically distinctive, *globally*, from all other slides in the slideshows.

The data from the chronological condition replicate those from the first experiment in showing that although change and looking time rise and fall in tandem, change does not completely account for looking time, further evidence that there are subtle or more conceptual changes that signal a transition that viewers detect but that the change measure does not. These more conceptual changes might relate to the particular objects being acted on and their associated goals. For activities like these, each new action unit at the coarse level typically entails a new object or object part as well as a new action, whereas at the fine level, each new action unit typically entails a new action on the same object or object part (Zacks, Tversky, & Iyer, 2001). Introducing both a new object and action likely explains some of the greater physical change seen at coarse breakpoints, compared with fine breakpoints. But physical change aside, viewers' recognition of a new object and the new encompassing goal that comes along with it could also affect attention.

Not surprisingly, participants who viewed chronological slideshows recalled more correct actions and made fewer intrusions than did participants who viewed the scrambled slideshows. As before, participants who looked longer at breakpoints recalled more correct actions. Thus, relative looking time is a good index of online mental organization of the ongoing events and better online organization yields better memory. Together, the results of this study corroborate and extend the previous results. Importantly, they show that breakpoint moments are intrinsically different from ordinary moments, so they are looked at longer than other moments, even out of temporal context. Breakpoint moments are even more privileged in context, in a coherent temporal structure.

General Discussion

For much of their lives, people are engulfed in an ever-changing stream of activity. Ultimately, people need to make sense of the chaotic confusion, to focus on, absorb, and act on the information that is important. In all the senses and in all domains, people cope with information overload by segmenting and organizing it, chunking it into kinds and parts. For the multisensory information dispersed over time, people form categories of events and actions, and organize those into taxonomies of kinds (e.g., Morris & Murphy, 1990) and partonomies of parts (e.g., Tversky et al., 2008; Zacks, Braver, et al., 2001; Zacks & Tversky, 2001). By segmenting the stream of activity, observers transform what is inherently continuous and unformed into units that are discrete and meaningful.

The breakpoints that segment the action stream into units are key to this transformation. Breakpoints have a privileged status in the perception and understanding of activities: There is agreement on their locations, they are hierarchically organized and are especially memorable and comprehensible, and deleting them is disruptive to understanding (e.g., Hard et al., 2006; Newtson & Enquist, 1976; Schwan & Garsoffky, 2004; Zacks, 2004; Zacks & Tversky, 2001). The work described here extended those findings in a number of directions by introducing two new methods.

Because breakpoints mark the convergence of important perceptual and conceptual cues, we predicted that breakpoints would be marked at local maxima of action change. More change means more information to process, which should elicit heightened attention. To assess heightened attention, we developed the *dwell time* procedure: videos of organized human activity were sampled at a 1-s rate and presented as a slideshow. Participants were free to look at each slide as long as they wished in order to remember the actions portrayed. To reveal the contour of action change, we computed the slide-to-slide, pixel-to-pixel differences between consecutive slides that were first filtered to highlight human movements. Later, participants segmented the source videos into coarser and finer units.

Although both measures are crude—dwell time does not reveal what is attended to and change does not specify what is changing—their joint time courses revealed much about how observers segment ongoing action. Breakpoints were marked at local maxima of action change and elicited greater attention, as indexed by dwell time. Moreover, both change and looking time were greater for coarser breakpoints than for finer breakpoints, confirming that observers segment ongoing behavior at multiple levels simultaneously. Increased attention to breakpoints indicates better hierarchical organization, which should improve memory. In fact, observers who looked relatively longer at breakpoints remembered more actions than those who looked less.

Although change and looking time were closely coupled, regression analyses showed that change did not completely account for looking time. More subtle changes in the flow of action may contribute to looking time, as well as conceptual changes, such as accomplishment of goals and subgoals. There were also hints of qualitative differences between physical change and looking time contours: Prior to breakpoints, looking time increased several slides before change increased. It is possible that observers attended to small yet meaningful changes such as the shifts of eye gaze and body that begin prior to a breakpoint and signal an impending transition from one action to another.

Is the informativeness of breakpoints entirely a consequence of being embedded in a sequence of actions, or is there also informativeness inherent in the breakpoints? The second experiment confirmed that even when the slide sequence was scrambled, breakpoint slides were more different from the preceding slides, and they attracted more looking time. That is, the privileged status of breakpoints is in part inherent; they are more distinctive than ordinary moments, even out of context. Notably, the informativeness of breakpoints increased in context. There was a stronger increase in both looking time and physical change at breakpoints for chronologically viewed sequences and clear differentiation in looking time for breakpoints at different levels of segmentation. Unlike for scrambled sequences, looking time for breakpoints in chronological sequences could not be fully explained by the physical distinctiveness of the breakpoint slides. These findings replicated the earlier findings, despite the breakpoints being determined by separate observers.

How these findings will extend to the observation of other types of activities remains to be seen. We explored an extremely modest range of activities designed to be familiar to undergraduates. What about activities that are unfamiliar? Given the convergence of bottom-up and top-down information across an activity, we can predict that unfamiliar activities would produce similar patterns of dwell time. The redundancy of cues to breakpoints mean that even when one kind of cue is unavailable, other kinds of cues can allow recognition and permit inference and action (e.g., Hard et al., 2006). What about activities that differ in organization? Activities vary in the extent to which they are hierarchically organized and in the constraints that are imposed on the relations between actions in a sequence. Some relations are causally constrained, for example, putting the bottom sheet on the bed before the top sheet. Others are not; the bills can be paid and the dishes washed in many orders. Whether and how these constraints affect the processing of activities across time remains an open question. Finally, the present experiments involved monitoring looking time to slices of an activity under instructions to study them for later recall. Certainly, people do not normally attend to activities for the purpose of memorizing them, and it would be informative to test other instructions.

To act effectively in the world, people need to make sense of what they see, and sense-making entails understanding actions as they unfold in the world. Understanding requires partitioning the action stream into meaningful units. The current project has shown that local maxima of action change draw attention and are an excellent cue to the critical moments of the action stream: the completion of one goal and the initiation of actions directed to another. The convergence of perceptual and conceptual cues at breakpoints, along with the allotment of added attention, should encourage inferring one cue from the other, promoting learning action sequences, understanding the intentions of others, and planning one's own actions.

Breakpoints are links. They link one action to another: the completion of one action and the simultaneous preparation to enact the next. They link the perceptual to the conceptual: from observing a movement to inferring its goal. They link one level of an action hierarchy to another: coarser to finer. Breakpoints are more than boundaries between actions; they mark the confluence of concrete and abstract features that provide the shape of ongoing action.

References

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

Baldwin, D. A., & Baird, J. A. (1999). Action analysis: A gateway to intentional inference. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 215–240). Mahwah, NJ: Erlbaum.

Barker, R. G., & Wright, H. F. (1955). *Midwest and its children.* Evanston, IL: Row Peterson.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–220. doi:10.1016/0010-0285(79)90009-4

Dickman, H. R. (1963). The perception of behavioral units. In R. G. Barker (Ed.), *The stream of behavior* (pp. 23–41). New York, NY: Appleton-Century-Crofts.

Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General, 118,* 136–147. doi:10.1037/0096-3445.118.2.136

Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34,* 1221–1235. doi:10.3758/BF03193267

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72–79. doi:10.1016/j.tics.2007.11.004

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–146). New York, NY: Wiley.

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research, 179,* 427–442. doi:10.1007/s00221-006-0804-0

Morris, M. W., & Murphy, G. L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition, 18,* 407–418. doi:10.3758/BF03197129

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28,* 28–38. doi:10.1037/h0035584

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

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35,* 847–862. doi:10.1037/0022-3514.35.12.847

Pierno, A. C., Becchio, C., Wall, M. B., Smith, A. T., Turella, L., & Castiello, U. (2006). When gaze turns into grasp. *Journal of Cognitive Neuroscience, 18,* 2130–2137. doi:10.1162/jocn.2006.18.12.2130

Radke, R. J., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing, 14,* 294–307. doi:10.1109/TIP.2004.838698

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science, 31,* 613–643. doi:10.1080/15326900701399913

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

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

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review, 69,* 344–354. doi:10.1037/h0043150

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193. doi:10.1037/0096-3445.113.2.169

Tversky, B., Zacks, J. M., & Hard, B. M. (2008). The structure of experience. In T. Shipley & J. Zacks (Eds.), *Understanding events: How humans see, represent, and act on events* (pp. 436–464). New York, NY: Oxford University Press.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28,* 979–1008. doi:10.1207/s15516709cog2806_5

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

Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movements and intentions to understand human activity. *Cognition, 112,* 201–216. doi:10.1016/j.cognition.2009.03.007

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind–brain perspective. *Psychological Bulletin, 133,* 273–293. doi:10.1037/0033-2909.133.2.273

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127,* 3–21. doi:10.1037/0033-2909.127.1.3

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

Appendix A

Calculation of Physical Change

First, each frame of the slideshow was passed through a convolution filter that identified and highlighted high contrast areas of the image, such as the edges of people and objects in each frame (see Figure A1). Each pixel of the resulting image was then assigned a numeric contrast value from 0 to 255, corresponding to its brightness (as defined by the hue–saturation–brightness color model). Contrast values were high for pixels near the edges of people and objects in each frame, but low elsewhere.

The next step in the calculation was to pair each pixel in that slide with the corresponding pixel in the previous slide, determine the absolute value of the difference between the contrast values for every such pair, and sum up all the resulting values to yield a single change value for that slide. This calculation can be formally defined as

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left| p_{ij} - q_{ij}\right|$$

where p_{ij} is the contrast value of the pixel at coordinate (i, j) in the first frame, where q_{ij} is the contrast value of the pixel at coordinate (i, j) in the second frame, and where n and m represent the width and height of the slideshow in pixels.

(*Appendices continue*)

Figure A1. An example of what an original slide (left) looked like after it was passed through a convolution filter (right).

Appendix B

Calculation of Alignment and Enclosure Scores

Alignment

Alignment is a measure of hierarchical segmentation that captures how temporally aligned the boundaries of coarser- and finer-level action units are. This continuous alignment measure was developed by Zacks, Tversky, and Iyer (2001). Alignment was first calculated between coarse and intermediate levels of segmentation. For each coarse breakpoint, the temporal distance in ms to the nearest intermediate breakpoint was calculated. These distances were averaged to determine the mean distance (AveDist) for each participant. The mean distance was compared with a null model—an *expected* distance between coarse and intermediate breakpoints based on chance (AveDist_0). Calculations of this null model are described in detail in Zacks, Tversky, and Iyer (2001). This method revealed that coarse and intermediate breakpoints were more aligned (M = 1,708.42 ms, SEM = 216.76) than chance (M = 6,400.05, SEM = 731.89), $t(39) = 6.84$, $d = 1.37$. Alignment was next determined between intermediate and fine breakpoints. Fine and intermediate breakpoints were more aligned (M = 911.30 ms, SEM = 72.48) than chance would predict (M = 2,645.34, SEM = 351.91), $t(39) = 5.41$, $d = 1.08$. On the basis of the reasoning that a larger difference between expected and observed alignment should indicate greater hierarchical encoding, we also calculated a set of alignment scores for each participant by subtracting the observed alignment from the alignment expected by chance and dividing that difference by the alignment expected by chance, following Hard et al. (2006).

Enclosure

Enclosure is a measure of hierarchical segmentation that captures how much coarser level action units contain, or enclose, finer level units. Intermediate–coarse enclosure was calculated according to the following steps. The left side of Figure B1 shows a chronological list of a set of coarse breakpoints and intermediate breakpoints. These numbers represent each time in ms, from the start of the video, that the participant pressed the spacebar to indicate a coarse and an intermediate breakpoint. The first step was to take each coarse breakpoint and line it up with the intermediate breakpoint that it was temporally closest to, as shown in the first two columns on the right side of the figure. Next, once a coarse breakpoint was lined up with an intermediate breakpoint, we determined whether that coarse breakpoint occurred temporally before or after the fine breakpoint, as shown in the final column of the figure.

The numerator of the enclosure score was then determined by first checking for cases in which multiple coarse breakpoints shared (i.e., were closest to) the same intermediate breakpoint. For each such case, we determined which of the shared coarse breakpoints was in fact closest to the intermediate breakpoint. Only this pairing would be used in determining the participant's enclosure score; the other pairing was excluded. One consequence of this rule is that the enclosure score is penalized when a coarse breakpoint does not reasonably line up with an intermediate breakpoint.

(*Appendices continue*)

Coarse Breakpoint Times	Intermediate Breakpoint Times
3781	3811
41085	5011
65108	9811
92300	17931
101676	23315
129684	32603
171651	38595
	45003
	45339
	51186
	55107
	59386
	77867
	91050
	99762
	115474
	128106
	142866
	150418
	166769
	167929
	170721
	173409
	175705
	176113
	177002
	177337
	178985
	180065
	180841

Coarse Breakpoint Times	Intermediate Breakpoint Times	Is coarse breakpoint before or after intermediate breakpoint?
3781	3811	Before
	5011	
	9811	
	17931	
	23315	
	32603	
41085	38595	After
	45003	
	45339	
	51186	
	55107	
65108	59386	After
	77867	
92300	91050	After
101676	99762	After
	115474	
129684	128106	After
	142866	
	150418	
	166769	
	167929	
171651	170721	After
	173409	
	175705	
	176113	
	177002	
	177337	
	178985	
	180065	
	180841	

Figure B1. An illustration of enclosure calculation between coarse and intermediate breakpoints. Lists of the participant's coarse and intermediate breakpoint times are shown on the left. The columns on the right depict the line up of coarse breakpoints with the nearest intermediate breakpoint and the determination of whether that coarse breakpoint fell before or after the intermediate breakpoint in time.

The numerator of the enclosure score is then equal to the remaining total number of cases in which a coarse breakpoint fell after its nearest intermediate breakpoint. In the example, the participant has an enclosure score numerator of 6.

The enclosure score was calculated by taking the numerator and dividing it by the total number of coarse units. In this example, the participant has a total of 7 coarse units, so we calculate the enclosure score to be 6/7 = .86. The maximum possible enclosure score would be 1, which would mean that all of the coarse breakpoints followed their closest intermediate breakpoint in time rather than preceding them. The minimum enclosure score would be 0, indicating that all of the course breakpoints preceded their closest intermediate breakpoint instead of following them. Fine-intermediate enclosure was similarly calculated by determining the proportion of intermediate breakpoints that followed their closest fine breakpoint in time.

Received September 22, 2010
Revision received April 25, 2011
Accepted May 2, 2011 ■