Attention reorganizes as structure is detected in dynamic action

Bridgette Martin Hard¹ · Meredith Meyer² · Dare Baldwin³

Published online: 30 August 2018
© Psychonomic Society, Inc. 2018

Abstract

Once one sees a pattern, it is challenging to “unsee” it; discovering structure alters processing. Precisely what changes as this happens is unclear, however. We probed this question by tracking changes in attention as viewers discovered statistical patterns within unfolding event sequences. We measured viewers’ “dwell times” (e.g., Hard, Recchia, & Tversky, 2011) as they advanced at their own pace through a series of still-frame images depicting a sequence of event segments (“actions”) that were discoverable only via sensitivity to statistical regularities among the component motion elements. “Knowledgeable” adults, who had had the opportunity to learn these statistical regularities prior to the slideshow viewing, displayed dwell-time patterns indicative of sensitivity to the statistically defined higher-level segmental structure; “naïve” adults, who lacked the opportunity for prior viewing, did not. These findings clarify that attention reorganizes in conjunction with statistically guided discovery of segmental structure within continuous human activity sequences. As patterns emerge in the mind, attention redistributes selectively to target boundary regions, perhaps because they represent highly informative junctures of “predictable unpredictability.”

Keywords Action perception · Event segmentation · Statistical learning · Dwell time

Most of us have had the experience of encountering an utterly unfamiliar language. Initially, what we hear seems an impenetrable wall of continuous babble, only to morph, with time and experience, into organized, meaningful linguistic units—words, phrases, utterances—that we can readily interpret. Something analogous occurs in the domain of action: A novel and continuous activity like knitting, tying shoes, or break-dancing initially seems baffling. Yet it can transform with experience into an orderly and interpretable sequence, with small, goal-oriented motions organized into larger-scale intentional acts. In both language and action, observers discover segmental structure within the dynamic stimulus stream, and discovering this structure changes how the stimulus stream is experienced in the future (e.g., Baldwin & Baird, 2001; Christiansen & Chater, 2016; Gobet et al., 2001; Koo, 2008; Pani, 2000; Saffran et al., 1996a).

Precisely *how* does the processing of a continuous stream change with experience? How does one’s discovery of segmental structure change the way that a novel action stream is interpreted and encoded? Investigating these questions requires methods for measuring learners’ processing of a novel action stream as it unfolds across time. Relevant techniques exist and have produced important findings in relation to language processing (Cutler, 2012; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009; Sanders, Ameral, & Sayles, 2009). In the domain of action processing, Hard and colleagues introduced a novel looking-time technique—since dubbed the “dwell-time paradigm”—that offers this capability (Hard, Recchia, & Tversky, 2011). A primary purpose of the present research was to deploy the dwell-time paradigm to examine observers’ emerging sensitivity to segmental structure as they experience an initially novel, unfolding activity sequence. We hypothesized that characteristic patterns of attentional change would arise in conjunction with viewers’ emerging segmental analysis of the sequence.

In the sections that follow, we first summarize key findings regarding patterns of attention for observed human action. We then offer an explanation for these patterns that led to the predictions of the present study. Finally, we propose an approach to testing these predictions that unites the dwell-time paradigm with another technique used to study the processing of dynamic stimuli: the statistical-learning paradigm.

✉ Bridgette Martin Hard
bridgette.hard@duke.edu

¹ Department of Psychology and Neuroscience, Duke University, Durham, NC 27708, USA

² Department of Psychology, Otterbein University, Westerville, OH, USA

³ Department of Psychology, University of Oregon, Eugene, OR, USA

Systematic dwell-time patterns: A boundary advantage in observed action

Hard et al.'s (2011) dwell-time technique probes viewers' attention as they advance through a series of still images (a "slideshow") depicting dynamic activity. Viewers watch slideshows that are constructed by extracting images ("slides") chronologically at a regular increment (e.g., once every 500 ms) from a digital video of natural activity. Viewers advance images at their own pace via a key press or mouse click, giving them flexibility in how long they attend to each image.

Although viewers move quickly through such slideshows, Hard and colleagues observed that the time spent viewing particular slides—called dwell times—reflected systematic patterns of attention. Viewers dwelled longer on slides depicting segment boundaries within the unfolding event, a pattern we call a *boundary advantage*. Furthermore, they dwelled longest on slides depicting higher-level "coarse-grained" boundaries (e.g., finishing emptying a dishwasher and moving on to sweeping the floor in a kitchen-cleaning event), relative to boundaries of component segments (such as completing the emptying of a glass or, even finer, completing the grasp of a glass). The boundary advantage in dwell time is consistent with other evidence for preferential attention to event boundaries. Studies have found that memory and perceptual detection are enhanced for relevant stimuli at event boundaries (Baldwin & Pederson, 2016; Swallow, Zacks, & Abrams, 2009) and weakened for irrelevant stimuli (Huff, Papenmeier, & Zacks, 2012). Together, these patterns point to dwell times indexing viewers' segmental analysis of unfolding event streams, as well as their representation of hierarchical relations among the segments they detect.

One interpretation of this package of findings is broadly consistent with ideas others have offered (e.g., Kurby & Zacks, 2008; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Viewers attend to segment boundaries (and especially to boundaries at higher levels of analysis) in part because these junctures are particularly informative regarding the structure and nature of ongoing activity, perhaps providing "conceptual bridges" in achieving an integrated understanding of that activity.

Explaining the boundary advantage: Attentional reorganization to learned segmental structure?

In what sense are segment boundaries informative? If, as in information theory, we define "information" in terms of low predictability (entropy; e.g., Hartley, 1928; Shannon, 1948), then a possible answer to this question is that segment boundaries are unpredictable for observers. That is, once an

unfolding action has been recognized, what immediately follows in the motion stream is generally predictable until that action is completed. But once the action is completed, a range of possible new actions could ensue, making the transition at the end of one action to the beginning of the next unpredictable. For example, once an actor grasps a coffee mug, a variety of possible actions might follow, such as lifting the mug, sliding the mug across the table, or tipping the mug over to one side. According to event segmentation theory (Zacks et al., 2007), this unpredictability is key to how event boundaries are detected in the first place.

However, to some extent, at least for a knowledgeable viewer, unpredictability is itself predictable, a point that is not yet incorporated into event segmentation theory. In other words, observers can tell, on the basis of their ongoing understanding of the activity, that an action is about to be completed and that the motion stream will soon become unpredictable. Consistent with this idea, dwell time tends to increase just before, rather than in response to, the completion of an action (Hard, et al., 2011). As an observer anticipates that an action is about to be completed, increasing attention and maintaining that attention until the next action begins would promote efficient processing of the unfolding activity stream.

If this interpretation is correct, then increases in dwell times at segment boundaries arise as a result of learning-driven modulation of selective attention. That is, as viewers gain familiarity with the structure underlying an activity sequence, their attention reorganizes. They increasingly direct their attention toward highly informative, "predictably unpredictable" regions such as segment boundaries (and perhaps, simultaneously, reduce their attention to less informative, midsegment portions of the motion stream). Conversely, viewers lacking familiarity with the structure of an activity sequence would have little basis on which to identify especially informative portions of the unfolding activity. Naïve viewers would thus be less likely to increase their attention at structurally relevant points within the activity sequence, such as segment boundaries.

These predictions of the attentional reorganization account seem particularly plausible, given other findings that sequence learning alters viewers', or listeners', attention (e.g., Astheimer & Sanders, 2011; Kidd, Piantadosi, & Aslin, 2012, 2014; Mayberry, Livesey, & Dux, 2010; Olson & Chun, 2001; Swallow & Zacks, 2008). For example, using a visual statistical-learning task involving sequences of novel abstract shapes, Turk-Browne, Jungé, and Scholl (2005) found that viewers who successfully learned the statistical patterns displayed an attentional pattern indicative of this newly learned statistical structure. Similar attentional reorganization processes may also be key to everyday event-processing phenomena, such as viewers' ability to selectively attend to boundary regions in dynamically unfolding activity sequences.

Considering alternatives to the attentional reorganization account

The attentional reorganization account poses that, as people discover structure in events, they reshape their attention toward that structure. This view may seem at odds with evidence that experience has little, if any effect, on how viewers detect structure in events. This has been studied by having observers press a key to dynamically mark boundaries, or “breakpoints,” between events in a video of naturalistic human, or sometimes animated, activity (Newton, 1973; Newton & Engquist, 1976). Several studies have revealed remarkable consistencies in segment length, organization, and even perceived boundary locations, regardless of observers’ prior knowledge—or even comprehension—of the activity (Hard, Tversky, & Lang, 2006; Zacks, Tversky, & Iyer, 2001).

These findings hint at an alternative interpretation of the boundary advantage in event processing. Event boundaries may garner attention because they coincide with a greater magnitude of changes in physical motion (e.g., Hard et al., 2006; Newton, Engquist, & Bois, 1977; Zacks, 2004; Zacks, Kumar, Abrams, & Mehta, 2009). The relationship between motion change and event boundaries provides a compelling explanation for why viewers can segment activity streams in remarkably similar ways, regardless of prior knowledge or understanding of the sequence: Bottom-up movement characteristics provide clues to event boundaries when top-down knowledge is absent. Hard et al. (2006) controlled for motion change in their analyses and found that it did not fully account for the boundary advantage, especially at increasingly coarse-grained levels of analysis. Nevertheless, their measure of physical change was crude and may not have captured all of the low-level features of event boundaries that influence attention.

Yet another account of the boundary advantage has remained untested: Given that action tends to decelerate at boundary regions (e.g., Zacks, 2004), viewers might slow their pace at segment boundaries during slideshow viewing in order to generate a visual experience—decelerating at boundary regions—that mirrors the temporal dynamics of everyday sensory experience. This “temporal mirroring” account stands as another alternative to the attentional reorganization account we have forwarded.

The familiar, naturalistic activity sequences utilized in prior dwell-time research were not amenable to disentangling these explanations for the boundary advantage in dwell times, because boundary regions in such sequences conflate predictable unpredictability with points of motion change and points at which human actions tend to decelerate.

Testing the attentional reorganization account

One route to testing the attentional reorganization account would be to present viewers with a novel motion stream involving an unfamiliar segmental structure that can only be discovered with experience. To rule out alternative accounts of the boundary advantage in dwell times, it is critical that the structure of the motion stream not be confounded with points of motion change or deceleration in motion. The central prediction is this: When dwelling through this activity sequence, viewers who have had an opportunity to discover its inherent segmental structure should display a boundary advantage indicative of this structure in their dwell times, whereas viewers without prior opportunity to discover such structure should not.

In the present research, we tested this prediction by uniting the dwell-time paradigm with the statistical-learning paradigm, a technique widely used to study the discovery of segmental structure in language (e.g., Saffran, 2003) and, more recently, in dynamic human action (Baldwin, Andersson, Saffran, & Meyer, 2008). Use of the statistical-learning paradigm in the action domain involves presenting observers with an extended, novel, meaningless sequence of small motion elements involving an actor and one or more objects (Baldwin et al., 2008). For example, the actor might *stack* a sponge on top of a glass bottle, *poke* her finger into the same bottle, and then lift the bottle to her lips as if to take a *drink*. Although the motions seem randomly ordered on first look, a segmental structure underlies the sequence: small motion elements are actually grouped into four distinct triplets (e.g., *stack–poke–drink*) that constitute higher-level action segments. Recent research has indicated that observers can detect this segmental structure and successfully discriminate between “actions” and foils that consisted of reordered motion element triplets (e.g., Baldwin et al., 2008; Meyer & Baldwin, 2011; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014).

The statistical-learning paradigm is well-suited to investigating unique predictions of the attentional reorganization account. It is a learning paradigm in which sensitivity to higher-level segmental structure is absent at first and is acquired only with exposure to the relevant statistical regularities; thus, it provides the opportunity to directly capture how processing changes as learning proceeds. It presents actions in which higher-level segmental structure is defined by underlying statistical regularities and is not confounded by deceleration or position changes. Thus, the paradigm provides an opportunity to rule out the motion change and temporal mirroring accounts of the boundary advantage pattern.

By uniting the statistical-learning paradigm with the dwell-time paradigm, we can discern the extent to which learners’ growing familiarity with statistically defined segmental

structure corresponds with systematic alterations in the way attention is allocated to the unfolding activity stream.

Overview of the present study

In the present study, we compared the dwell-time patterns of (a) viewers who had prior opportunity to learn the statistically defined higher-level segmental structure through viewing of an exposure corpus (“knowledgeable” viewers) and (b) viewers who had no such prior learning opportunity (“naïve” viewers).

First, we aimed to replicate prior findings that viewers can discover segmental structure in action on the basis of statistical regularities. Second, we aimed to extend prior research using the dwell-time paradigm by testing a two-part prediction of the attentional reorganization account. We predicted that observers would easily perceive boundaries between segments at the small-motion-element level (e.g., *stack* vs. *poke* vs. *drink*), on the basis of their prior understanding of these simple actions. Thus, we predicted that participants would display a boundary advantage for lower-level segmental structure within the activity stream, regardless of their prior viewing experience. In contrast, we predicted that statistical learning would be required in order to perceive boundaries between higher-level “actions” in the sequence. Knowledgeable and naïve participants differed in their opportunities to learn this statistically defined, higher-level segmental structure. Thus, we predicted that only the knowledgeable group would display a boundary advantage in their dwell times at the higher level of analysis.

Finally, we explored the relationship between dwell-time patterns and viewers’ explicit knowledge of action structure. We wondered whether participants would show a boundary advantage for high-level action segments only when they could also display knowledge of these segments in a discrimination judgment task, or whether dwell-time patterns might reveal implicit awareness of segmental structure, even when participants couldn’t reliably express that knowledge in their discrimination judgments.

Method

Participants

We selected our sample size by examining prior published work on the statistical learning of novel action sequences (e.g., Baldwin et al., 2008; $N = 17$ to 24 per study) and also on dwell-time patterns (e.g., Hard et al., 2011; $N = 40$ per study). We selected a larger sample size than had been used in either of these published works, to ensure adequate power for looking at the relationships between statistical-learning

measures and dwell-time patterns. A total of 64 university students (42 females, 22 males) were included in the final sample and received class credit in exchange for participation. Four additional participants completed the study but were excluded from analysis: three because of participating in similar studies of statistical learning, and one because of a computer malfunction during the task.

Materials

Creation of exposure materials and slideshows Slideshows were created from the exposure corpora originally used in studies by Baldwin et al. (2008) and Meyer and Baldwin (2011) to study statistical learning in the action domain. To examine the robustness of any effects, we selected two different, 11-min stimulus sets that each met the design criteria but varied in complexity. Both exposure corpora consisted of a set of films depicting 12 small motion elements (SMEs). In the first corpus, each SME was performed by a female actor on a glass bottle (used also in Meyer & Baldwin, 2011, Study 1), and each instance of the SME was identical. We called this corpus the *uniform* set, because of the use of a single object and exemplar for each motion pattern. The second exposure corpus included *categories* of similar, but not identical, instances of SMEs (used also in Baldwin et al., 2008, Study 3). This stimulus set featured the same actress manipulating a glass bottle, a sponge, or a drinking glass, and featured 28 different exemplars of each SME. We call this set the *variable* set, for the variable objects and exemplars of each motion pattern.

Within each set (uniform and variable), SMEs were grouped into four *actions*, with each action consisting of three randomly selected SMEs. Note that although the SMEs themselves were meaningful in the sense that their small-scale goals could likely easily be inferred (e.g., pour, clean, inspect, etc.), the higher-level actions they comprised (e.g., *pour-clean-inspect*) were arbitrary and without obvious intentional content. Next, the four actions for each set were randomly ordered into a continuous sequence. Figure 1 shows sample images from each corpus, and the Appendix Tables 1 and 2 provides additional details, including a list of the SMEs and their groupings into actions.

The two exposure corpora presented different challenges for learning the statistical regularities in the sequences. First, observers saw 40 tokens of each action in the uniform set and only 28 tokens of each action in the variable set, due to differences in the speeds of the SMEs in the two sets (see the Appendix for details). The higher number of tokens in the uniform set might have made it easier to learn. But because the uniform set involved SMEs performed on a single object, it also prevented observers from tracking different objects (e.g., bottle, glass, sponge) in order to help learn the sequence. The variable set involved SMEs performed with three

Fig. 1 Two sample small motion elements (SMEs) from the uniform set (left) and the variable set (right). Each set included a total of 12 different SMEs. The uniform set showed only one exemplar of each SME, but the variable set showed 28 exemplars of each (the figure shows just two exemplars).

different objects, although most of the actions included at least two SMEs performed on the same object (i.e., a bottle). The variable set challenged observers to track the categories of SMEs, rather than single exemplars. Because there were features of each exposure set that made that set challenging to learn, we thought it possible that the learnabilities of the sequences might differ, though we had no specific predictions. Our primary interest was whether our hypothesized findings would hold across two sets of previously studied stimuli.

Once the exposure corpora were created, two separate “slideshows” were constructed by extracting still-frame images from each of the two exposure corpora (uniform and variable). Pilot testing had indicated that, given the speed of the SMEs, an extraction rate of four images/second was the lowest that would make each SME discernable to viewers. Each slideshow contained approximately 2,800 images.

Creation of test materials For each exposure corpus, video clips of each of the four actions were created for use in the test phase. We also created sequences for use at test called *part-actions*, which were sequences of three SMEs that spanned action boundaries (see Fig. 2 for an example). Each trial in the test phase required discrimination between a pair of SME triads: an action and a part-action. We exhaustively paired actions with part-actions to create 16 test trials. The orders of action versus part-action presentation within a single test trial, as well as the test trials themselves, were random. The Appendix Tables 1 and 2 provides additional details regarding the test stimuli.

Procedure

Participants in the knowledgeable condition ($n = 32$) first experienced an exposure phase in which they watched the exposure corpus video, showing a stream of dynamic action (either uniform or variable version). Second, they completed the dwell-time phase, in which they advanced themselves through the still

images in the corresponding slideshow by pressing the right arrow key on the keyboard. Participants could only advance slides forward and were told they would answer questions about what they saw after the viewing. We did not provide any further instructions. Following the action exposure video and slideshow, participants completed the test phase, in which they were asked to select which SME triad (action vs. part-action) was more familiar to them on the basis of their previous exposure to the activity sequence. The participants in the naïve condition ($n = 32$) performed the identical task, except without the exposure phase: They did not watch the exposure corpus video prior to the slideshow, but only participated in the dwell-time phase and the test phase. A script in E-Prime recorded participants’ key presses during the slideshow and their forced-choice selection of actions versus part-actions during the test phase.

Dwell-time measure

Dwell time corresponds to how long a given slide was visible on the computer screen, via the latency between key presses during the slideshow task. Following procedures defined by

Fig. 2 During the test phase, participants were asked to discriminate between an action they had seen in the exposure stream (e.g., pour–clean–inspect or touch–blow–peek) and a part-action—a triad that spanned an action boundary (e.g., clean–inspect–touch)

Hard et al. (2011) and Kosie and Baldwin (2018a), the dwell times taken from the first five slides were excluded from the analysis, as were dwell times more than three standard deviations from the overall group mean. The data were positively skewed, as is typical for looking-time data. A $\log_{10}$ transformation was used to normalize the distribution prior to analysis. As in other research, dwell times followed a power function, such that viewers increasingly reduced their dwell times as they progressed through the slideshow. A power function significantly fit the dwell-time data for 88% of the participants, accounting for an average of 16% of the overall variance in dwell times ($SD = 0.17$), with no significant differences across conditions. Thus, the data were next transformed by fitting a power function to the data (using the curve estimation function in SPSS) for each individual participant and calculating residuals in relation to these power curves. The resulting dwell times could thus be positive (above the curve) or negative (below the curve), with higher dwell times indicating longer looking.¹

Results

Overview of predictions

In the present study, observers viewed a novel activity stream designed with two levels of segmental structure: a lower level corresponding to the individual SMEs, and a higher level corresponding to groupings of three SMEs into “actions.” Actions were discoverable only by detecting the statistical regularities among the motion elements. We aimed to answer three primary questions regarding observers’ processing of this activity stream.

First, could observers discover the higher-level segmental structure with sufficient experience, replicating prior findings (e.g., Baldwin et al., 2008; Meyer & Baldwin, 2011)? Second, would observers show a boundary advantage, such that their dwell times would peak at slides corresponding to perceived boundaries in the activity stream (e.g., Hard et al., 2011)? We expected that participants would show a boundary advantage for the lower level of segmental structure regardless of their prior viewing experience (i.e., naïve as well as knowledgeable viewers), but that only participants with sufficient opportunity to detect the statistical regularities in the activity (i.e., knowledgeable viewers) would show a boundary advantage at the higher level, suggesting attentional reorganization. Third, what was the nature of the relationship between dwell time and explicit knowledge of higher-level segmental structure? To explore this question, we examined the extent to which a

boundary advantage in dwell-time patterns was related to accuracy in the discrimination test.

For all analyses, the criterion for significance was an alpha level of .05, and we report 95% confidence intervals of the difference. 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 .

Confirming the successful discovery of segmental structure

Our first analysis addressed whether viewers discovered the higher-level segmental structure in the novel activity sequences via sensitivity to statistical structure, replicating prior findings (e.g., Baldwin et al., 2008). We examined participants’ success at discriminating actions from part-action foils, and whether this success depended on prior familiarity with the activity sequences. Participants in the knowledgeable condition watched both a video and a slideshow of the activity, so that the total length of exposure to the statistical regularities roughly matched what had occurred in previous studies (Baldwin et al., 2008). Participants in the naïve condition viewed only the slideshow of the activity, and thus received only half the exposure that prior research had shown led to learning of the higher-level segmental structure within the activity stream.

Following procedures established by past studies of statistical learning (e.g., Saffran et al., 1996b), we analyzed the percentages of test trials for which participants reported actions to be more familiar than part-actions. A 2 (Experience: naïve vs. knowledgeable) $\times$ 2 (Exposure Set: uniform vs. variable) between-subjects ANOVA revealed a significant main effect of experience, $F(1, 60) = 6.30$, $p < .05$, $\eta_p^2 = .10$, 95% CI $[-21.7\%, -2.4\%]$, with enhanced performance in the knowledgeable condition ($M = 63.38\%$, $SD = 21.82$) relative to the naïve condition ($M = 51.31\%$, $SD = 15.75$). As predicted, and replicating prior findings, discrimination performance was better than chance for the knowledgeable condition, one-sample $t(31) = 3.47$, $p < .01$, Cohen’s $d = 1.25$, 95% CI $[5.5\%, 21.2\%]$, but not for the naïve condition, one-sample $t(31) = 0.47$, $p = .64$, Cohen’s $d = 0.17$, 95% CI $[-4.4\%, 7.0\%]$.

Recall that there were many reasons to predict that the segmental structure of one stimulus set might be easier to learn than that of the other; however, participants discriminated actions from part-actions equivalently well across the two sets (uniform set $M = 55.63\%$, $SD = 19.24$; variable set $M = 59.06\%$, $SD = 20.57$), $F(1, 60) = 0.51$, $p = .48$, $\eta_p^2 = .01$, 95% CI $[-13.1\%, 6.2\%]$. Additionally, no significant experience by stimulus set interaction emerged in the ANOVA ($p = .67$), indicating that the enhanced discrimination performance in the knowledgeable versus the naïve condition emerged to comparable degrees in both the uniform and variable versions of the activity sequence. We nevertheless examined

¹ In the analyses that follow, we included data from participants whose dwell times failed to show a significant power function. However, we reran the analyses excluding these participants and found that the pattern and significance of the results were consistent, regardless of whether these participants were included.

discrimination performance in the knowledgeable condition ($n = 32$) separately within the two versions, to confirm above-chance performance in both. Separate one-sample t tests indeed revealed better-than-chance identification of actions relative to part-actions in the knowledgeable condition for both the uniform set, $M = 62.69\%$, $SD = 20.45$, $t(15) = 2.48$, $p < .05$, and the variable set, $M = 64.06\%$, $SD = 23.76$, $t(15) = 2.36$, $p < .05$.

In sum, the first set of analyses provided clear replication of prior findings that viewers readily acquire higher-level segmental structure within novel activity sequences via the ability to track statistical regularities within those sequences.

Confirming a “boundary advantage” in dwell times

Second, we queried whether dwell-time patterns would replicate prior findings that viewers look longer at natural-unit boundaries, or “breakpoints,” relative to slides depicting within-unit content (e.g., Hard et al., 2011). We also investigated the extent to which experience with the activity sequence affected the boundary-advantage pattern, specifically for higher-level action boundaries.

Before performing these analyses, we first considered that dwell times reflect residuals off a power curve fitted to participants’ log-transformed looking times. The residualization transformation might mask any between-subjects differences in total dwell time. Thus, we compared the untransformed (absolute) looking times, in seconds, in a 2 (Condition: naïve vs. knowledgeable) $\times 2$ (Exposure Set: uniform vs. variable) between-subjects ANOVA, and found that total looking times at the slideshows did not vary as a function of either experience (naïve vs. knowledgeable), $F(1, 60) = 2.00$, $p = .16$, $\eta_p^2 = .03$, or exposure set (uniform vs. variable), $F(1, 60) = .003$, $p = .95$, $\eta_p^2 < .001$. On average, participants spent about 10.5 min ($SD = 3.2$) pacing through the slideshows.

Confirming that dwell times reflected sensitivity to lower-level segmental structure We predicted that all observers, regardless of experience with the activity sequence, would show a boundary advantage for lower-level segmental structure—namely, the boundaries between SMEs in the activity sequence. In prior findings using everyday intentional activity sequences, dwell times tend to increase for slides depicting lower-level action boundaries (as compared to nonboundary slide content), and to increase even more for higher-level boundaries (Hard et al., 2011). In the present activity sequence, the SMEs consisted of familiar and readily recognizable actions with natural movement characteristics, such as changes in the actor’s body position and trajectory that mark their completion. Thus, observers should need no special experience to identify the SME boundaries.

Before testing this prediction, a first step was to identify where boundaries were within the novel activity sequences employed in this research. Recall that each SME began and ended with the actor in the identical, neutral position, so that the elements could be sequenced in any order while maintaining a natural flow of motion. These neutral transitions might seem like obvious boundaries, but as compared to the boundaries in naturalistic action, which tend to correspond to points of goal completion, these neutral transitions were artificial and not informative as to the nature of the actor’s goals. For the SMEs, the more naturalistic action boundaries were the informative points at which goals were completed—such as the point when the bottle was fully elevated and the actor looking inside (*peek*) or when it was tilted to its maximum extent on taking a drink (*drink*) (for an example, see the bottom panels of Fig. 3). From this naturalistic perspective, action boundaries in the present stimuli were defined as points when the actor reached goal completion for each SME. Of interest, then, was whether viewers dwelled longer on slides depicting such boundaries than on slides depicting “within-element” content.

A trained coder visually inspected and classified the slides as depicting either boundary or within-element content, on the basis of the guideline to identify a slide as a boundary if “it represents the point of maximum extent—the point at which the main ‘peak’ of the action is before the actor starts to return to her starting position.” All slides not judged to be boundaries were then considered within-element slides. A second trained coder independently classified a subset of the slides (47%, or 2,483/5,288 slides, selected from the uniform version) in order to calculate the reliability of the boundary versus within-element judgments, which was excellent, $\kappa_{ppa} = .80$.

To determine whether the dwell time was maximal at lower-level segmental boundaries (predicted for both knowledgeable and naïve viewers), we calculated the average dwell times for boundary slides, for the one to five slides presented before the boundary (preboundary), for the one to five slides presented after the boundary (postboundary), and for the slide representing the neutral resting position between the motion elements. Recall that the dwell times were transformed using $\log_{10}$ and residualization procedures, as described in the Method section. We then conducted a 4 (Slide Type: preboundary, boundary, postboundary, neutral) $\times 2$ (Experience: naïve vs. knowledgeable) mixed-design ANOVA examining mean dwell times, with slide type as a within-subjects factor and experience as a between-subjects factor. The data were collapsed across exposure sets (uniform vs. variable) because preliminary analyses had revealed no effects of exposure set on the dwell-time patterns nor any interaction with other variables. We observed a main effect of slide type, $F(3, 186) = 4.88$, $p < .01$, $\eta_p^2 = .07$. As is shown in Fig. 3, the dwell time increased and was maximal at boundary slides. Pairwise comparisons revealed that the dwell time for boundaries was significantly higher than those for slides

Fig. 3 Mean $\log_{10}$ -transformed and residualized dwell times as a function of lower-level segmental structure. Error bars represent standard errors of the means.

leading up to the boundary, 95% CI [.001, .005], $p < .01$; following the boundary, 95% CI [.002, .006], $p < .01$; and at the neutral resting point, 95% CI [.001, .007], $p < .01$, which did not differ significantly from each other. There was no main effect of experience, $F(1, 62) < 1$, on dwell times. Visual inspection of Fig. 3 hints at a sharper boundary advantage for knowledgeable than for naïve participants, but the interaction between experience and slide type was not significant, $F(3, 186) = 1.75$, $p = .16$, $\eta_p^2 = .03$.

These findings confirm a “boundary advantage” for the low-level segmental structure within these stimuli that was comparable to that seen in more naturalistic human action events, with dwell times peaking at information-rich boundaries within the motion stream (e.g., Hard et al., 2011). This boundary advantage at the low level of segmental structure was not affected by how much experience participants had with the exposure corpora.

Confirming that dwell times reflected acquired sensitivity to the higher-level segmental structure Next we focused on dwell times as an index of acquired sensitivity to higher-level segmental structure in the unfolding event stream. In prior research with naturalistic action sequences, dwell times showed sensitivity to higher-level segmental structure, in that viewers dwelled longer on boundary slides at higher (coarse-grained) than at lower (fine-grained) levels. Given the novel, statistically defined nonsense “actions” utilized in the present

research, how would dwell times reveal sensitivity to higher-level actions? The dwell time at segment boundaries is arguably driven by the need to increase attention toward predictably unpredictable, and thus highly informative, parts of the activity stream. For observers who have acquired the segmental structure of the continuous sequence, the activity stream is predictably unpredictable at the higher-level boundary, which is the onset of the first motion element of each action triad. Once knowledgeable viewers had identified the first motion element of an action, the remaining motion elements within the action should be increasingly predictable, and so require less attention. Thus, we predicted that knowledgeable viewers would display a triad-based linear trend in their viewing, such that dwelling would be longest for slides from the first motion element within a given action (e.g., on *stack* from *stack-poke-drink*) and decreasing for slides across the second and third motion elements (e.g., *poke* and *drink* from *stack-poke-drink*). We expected that naïve viewers would not show this triad-based linear pattern in dwell times.

We tested this prediction by first determining the mean dwell times for all slides in each of the first, second, and third SMEs within each “action,” excluding slides representing the neutral resting points. We then conducted a 2 (Experience: knowledgeable vs. naïve) $\times$ 3 (SME Type: first vs. second vs. third) mixed-design ANOVA, with experience as a between-subjects factor and motion element as a within-subjects factor. Overall, there was a main effect of motion

element type, $F(2, 124) = 17.87, p < .001, \eta_p^2 = .22$, such that, as predicted, the mean dwell time was highest for the first SME ($M = .0068, SD = .0107$), followed by the second ($M = .0050, SD = .0096$), and then by the third ($M = -.0013, SD = .0087$). Also as predicted, however, this effect interacted with experience, $F(2, 124) = 4.56, p < .05, \eta_p^2 = .07$. We analyzed the two experience conditions separately (repeated measures ANOVA with SME type as a within-subjects factor) with polynomial contrasts, to test for the presence of the triad-based linear trend. This analysis revealed different patterns of looking across the three motion elements, based on experience. For the knowledgeable group, the significant effect of motion element type on looking times, $F(2, 62) = 10.39, p < .001, \eta_p^2 = .25$, was best characterized by the predicted triad-based linear trend, $F(1, 31) = 13.14, p < .01, \eta_p^2 = .29$. The naïve group also showed a significant effect of SME type on looking times, $F(1, 31) = 13.65, p < .001, \eta_p^2 = .31$, but this effect was characterized by a quadratic trend, $F(1, 31) = 10.50, p < .01, \eta_p^2 = .25$, as well as by a linear trend, $F(1, 31) = 21.65, p < .001, \eta_p^2 = .41$. As is shown in Fig. 4, the mean dwell times in the knowledgeable condition followed a consistent downward trajectory from one SME to the next within each action unit. In the naïve condition, the mean dwell times were similar between the first and second SMEs, but then decreased to the third. Together, these findings suggest that dwell times were sensitive to observers' acquired knowledge of the segmental structure within the unfolding activity, such that observers who had an extended opportunity to acquire the higher-level statistical structure of an activity sequence prioritized their attention toward the predictably unpredictable higher-level boundary regions of the action stream. That is, they attended at highest levels specifically to the earliest portion of each action unit.

The overall time course of dwell times followed both low-level action and acquired high-level structure

Previous analyses had revealed that dwell times were sensitive to both low-level and high-level (acquired) action structure. Figure 5 summarizes how these effects unfolded across time.

Both naïve and knowledgeable observers showed a boundary advantage for low-level action structure: Their dwell times peaked when the goal of each SME was complete. Participants who had sufficient opportunity to discover the statistical structure governing how SMEs combined into larger “action” triads also displayed a clear boundary advantage for high-level action structure. Knowledgeable participants, who showed significantly higher accuracy in their explicit discrimination judgments, prioritized attention to the most informative and predictably least predictable parts of each action. Naïve participants, who had not had prior opportunity to discover the action structure, did not prioritize their attention in this efficient way, although they did show regularity in their dwell

time patterns. Naïve participants did decrease in dwell time from the second to the third motion element of each action triad, suggesting a budding implicit understanding of the statistical structure. We will examine this possibility further in the discussion.

Exploratory analyses: Were dwell-time patterns more sensitive to learning than explicit discrimination judgments?

An additional question of interest was the extent to which dwell times and the discrimination task indexed the same, versus divergent, knowledge. Data from the knowledgeable group provided a means of exploring this question. Although the participants in the knowledgeable condition could discriminate actions from part-actions on average, some did so more accurately than others. Given that implicit knowledge should give rise to explicit knowledge, more efficient patterns of dwell times should be correlated with recognition accuracy. This prediction was somewhat confirmed. Among knowledgeable participants, higher recognition accuracy was not significantly correlated with dwell times for the first SME, $r(32) = .19, p = .31$, or the second, $r(32) = -.24, p = .19$, but was significantly correlated with the dwell time for the third SME, $r(32) = -.37, p < .05$. Thus, more accurate participants tended to display briefer dwell times to the third, most predictable motion element of each action.

Does dwell time index implicit knowledge that the explicit discrimination task does not capture? To find out, we compared knowledgeable participants who displayed relatively strong discrimination accuracy ($n = 13$), by correctly discriminating more than 69%, the mean number of pairs (i.e., 11 out of the 16 items), to participants whose performance fell at or below the mean ($n = 19$). We then conducted a 2 (Learning Accuracy: high vs. low) $\times$ 3 (SME Type: first vs. second vs. third) ANOVA on dwell times, with learning accuracy as a between-subjects factor and SME type as a within-subjects factor. Consistent with the earlier analyses, we found a main effect of SME type, $F(2, 60) = 12.02, p < .001, \eta_p^2 = .29$, and no overall effect of learning accuracy, $F(1, 30) < 1$. If dwell times and explicit discrimination judgments were equally sensitive to participants' knowledge of higher-level segmental structure, we would expect that SME type would have interacted with learning accuracy, with the high-accuracy participants showing stronger dwell-time sensitivity to the higher-level segmental structure than low-accuracy participants. However, this was not the case, $F(2, 60) = 2.04, p = .14, \eta_p^2 = .06$.

As visual inspection of Fig. 6 shows, the triad-based linear trend appears steeper for the more accurate participants, yet the less accurate participants nevertheless clearly displayed a pattern suggesting implicit knowledge of the higher-level

Fig. 4 Mean $\log_{10}$ -transformed and residualized dwell times across the first, second, and third small motion elements within action triads, in relation to participants' prior knowledge of the statistical regularities. Error bars represent standard errors of the means.

action structure: They tended to dwell longer on the first SME in each action, followed by the second, and then the third. Given the small sample size, these exploratory findings should be interpreted with caution. They hint, however, that dwell-time patterns may be a more sensitive index of observers' knowledge of action structure than is the more standard, explicit discrimination task.

We also wondered whether the participants in the naïve condition who displayed relatively strong performance on the discrimination task might show a pattern of dwell times that was consistent with this knowledge. Unfortunately, only three out of the 32 naïve participants met the same accuracy

criterion (69% correct) applied to the knowledgeable group, making further comparisons impractical. A different approach to testing the sensitivity of dwell-time patterns to learning was to examine whether the dwell-time patterns in the naïve group changed across the duration of the slideshow, reflecting more of a triad-based linear trend in the second half, once participants had had more time to detect the underlying structure of the motion sequence.

To find out, we analyzed the dwell times of the 32 participants in the naïve condition with a 2 (Slideshow Half: first vs. second) by 3 (SME Type: first vs. second vs. third) ANOVA, with both slideshow half and SME type as within-subjects

Fig. 5 Summary of changes in $\log_{10}$ -transformed and residualized dwell times as a function of lower-level and higher-level segmental structures.

Fig. 6 Mean $\log_{10}$ -transformed and residualized dwell times in the knowledgeable condition as a function of acquired higher-level segmental structure and whether participants later showed high or low discrimination accuracy for the action triads. Error bars represent standard errors of the means.

factors. The analysis revealed the main effect of SME type previously reported, no main effect of slideshow half, $F(1, 62) < 1$, and a marginally significant interaction between slideshow half and SME type, $F(2, 62) = 2.69$, $p = .08$, $\eta_p^2 = .08$. Separate analyses of the first and second halves of the slideshow (repeated measures ANOVA with SME type as a within-subjects factor) with polynomial contrasts revealed subtly different patterns of dwell times for each half of the slideshow, as is shown in Fig. 7. The dwell times in the first half of the slideshow showed both significant linear, $F(1, 31) = 11.66$, $\eta_p^2 = .27$, and quadratic, $F(1, 31) = 18.93$, $\eta_p^2 = .39$, trends. In the second half of the slideshow, polynomial contrasts revealed a linear, $F(1, 31) = 6.66$, $\eta_p^2 = .18$, but not a quadratic, $F(1, 31) = 2.07$, $\eta_p^2 = .06$, trend. As Fig. 7 shows, the dwell times in the second half of the slideshow looked more similar to the triad-based linear pattern displayed by the knowledgeable group, although not as pronounced. All in all, then, naïve-condition participants tended to increasingly reveal a pattern of dwelling indicative of discovery of the higher-level segmental structure as their exposure to the relevant segment-related statistics accrued across time. These findings provided yet further evidence that attentional patterns reorganize as structure is discovered in novel activity sequences.

Discussion

This research introduced an innovative marriage of the dwell-time paradigm with the statistical-learning paradigm, to discover ways that viewers' attention reorganizes as they discover higher-level segmental structure within novel activity sequences. Two main findings emerged. First, the present results replicated findings from prior research employing these two methodologies. Regarding

statistical learning, we confirmed Baldwin and colleagues' (e.g., Baldwin et al., 2008; Meyer & Baldwin, 2011) finding that viewers can discover higher-level segmental structure within novel activity sequences via statistical learning. Regarding dwell-time patterns, we also confirmed replication of a characteristic boundary advantage—initially documented by Hard et al. (2011)—in viewers' dwell times as they advanced through a slideshow depicting the novel activity sequence.

Second, as predicted, this boundary advantage emerged as a function of experience. Low-level segmental structure—the level of the small motion element (e.g., *stack* vs. *poke* vs. *drink*)—was readily available to all participants, regardless of prior exposure to the novel action stream. For this low level of segmental structure, participants in both the knowledgeable and naïve conditions dwelled longer on slides depicting boundaries relative to within-segment slides. In contrast, the higher level of segmental structure was generally available to participants only after a significant degree of exposure. For this higher level of structure, knowledgeable participants' dwell times clearly revealed a boundary advantage—specifically, a triad-based linear trend: dwelling longest on slides depicting the first motion element of each action triad, with increasingly shorter dwell times to the second and third motion-element slides. As might be expected, naïve participants showed an emerging triad-based linear trend as they proceeded through the slideshow viewing; importantly, however, this pattern never reached the strength of the pattern for the knowledgeable group. The latter finding represents the first evidence to date that dwell times index the emergence of higher-level segmental structure as viewers discover that structure within continuously unfolding activity.

Fig. 7 Mean $\log_{10}$ -transformed and residualized dwell times in the naïve condition as a function of small motion element type and first versus second half of the slideshow. Error bars represent standard errors of the means.

Broader implications of the present findings

The present findings support the attentional reorganization account of the boundary advantage that has been repeatedly observed in viewers' dwell times—namely, that viewers dwell longer at boundaries in actions because these boundaries represent information-rich junctures that are high in predictable unpredictability. The present findings also rule out an alternative account of the boundary advantage that had remained untested: According to the temporal-mirroring account, viewers slow their pace during slideshow viewing in order to generate a visual experience that mirrors the deceleration naturally found at event boundaries. The familiar, naturalistic activity sequences utilized in prior dwell-time research conflated predictable unpredictability with points at which human actions tend to decelerate. The use of novel activity sequences, in which higher-level segments were defined solely by statistical regularities among the small motion elements, provided a unique opportunity to rule out the temporal-mirroring account of the boundary advantage pattern. The temporal-mirroring account predicts that dwell-time patterns should not change as viewers' knowledge of predictability structure grows. Consonant with the attentional reorganization account, dwell times indicative of higher-level segment discovery emerged only among knowledgeable viewers once they had acquired familiarity with the predictability structure of the activity sequence.

This same logic dismisses position change as the sole explanation for the boundary advantage. Hard et al. (2011) carefully examined the possibility that viewers dwell longer at event boundaries because the actor's position changes to a greater degree from slide to slide at boundaries than at slides within segments. In the activity sequences they utilized, increases in position change indeed coincided with boundaries (and to greater degrees at higher levels of segmental structure). However, a boundary advantage in dwell times persisted even

when position change was controlled for in their analyses, suggesting that position change was not the sole source of enhanced attention at boundaries. The present findings provide additional support for this conclusion. The observed differences in dwell-time patterns for knowledgeable versus naïve viewers suggest that position change alone is not an adequate account of the boundary advantage, at least at higher levels of segmental structure. Dwell-time changes also appear to reflect viewers' knowledge of the predictability structure within activity streams.

Two other aspects of the findings are especially noteworthy. First, our analyses revealed reassuring concordance across two methods for measuring viewers' learning of segmental structure in a novel activity sequence. Exposure to statistical regularities affected both the ability to perform the explicit discrimination task (the standard task in the statistical-learning literature for measuring acquisition of segmental structure) and dwell times (via the predicted triad-based linear trend in dwell times). Second, initial evidence emerged that dwell-time patterns appear to be more sensitive to viewers' acquired knowledge of segmental structure than has been apparent in the standard discrimination judgment task. This was revealed by exploratory analyses focused on knowledgeable participants' data. Knowledgeable participants displayed the expected triad-based linear trend in their dwell times even when their accuracy on the standard discrimination task was low, indicating implicit knowledge of the higher-level segmental structure. This exploratory finding needs to be replicated with a larger sample, but it hints that dwell time may tap emerging knowledge of segmental structure more sensitively than does a discrimination task. These findings mesh nicely with a range of other findings (e.g., Perruchet & Pacton, 2006) involving both language (e.g., Batterink, Reber, Neville, & Paller, 2015) and visual (e.g., Kim, Seitz, Feenstra, & Shams, 2009; Swallow & Zacks, 2008; Turk-Browne et al., 2005; Turk-Browne, Scholl, Chun, & Johnson, 2009)

sequence learning, indicating that implicit measures (e.g., reaction time and/or neuroimaging techniques) can provide information about learning that explicit tasks (e.g., discrimination judgments) fail to reveal. The present findings document that the dwell-time paradigm offers a useful addition to the available arsenal of tools for investigating event processing.

Caveats and limitations

The use of novel activity sequences offered a unique opportunity to rule out alternative accounts of the boundary advantage in dwell times—specifically, the temporal-mirroring and position change accounts. However, the use of novel activity sequences creates at least three limitations on the generalizability of the present findings. First, our stimuli exhibited a simplified statistical structure relative to everyday human activity. It may be considerably more difficult for viewers to discover segmental structure within naturalistic, everyday activity than it was for the participants in our research. Second, both the dwell-time paradigm and the discrimination task measured learning using actions identical to those viewed during the exposure phase. But event learning does not work in this way in real life—unless we are watching a recorded video, we rarely observe the same event in precisely the same way. Future work might test whether observers with the opportunity to learn segmental structure could extend their newfound knowledge to an action stream that differed more dramatically from the original, varying in actors, objects, and timings.

Third, in our novel sequences, the actor was in precisely the same neutral position at both the beginning and end of every small motion element, making it possible to combine any given small motion element with any other small motion element. Although such a return-to-neutral pattern exists in naturalistic activity (as when a kicker's leg returns to neutral when a kick is completed), everyday activity does not exhibit this property at the high rates necessitated by the statistical-learning design. This departure from ecological validity raises a different kind of concern about the generalizability of the findings. Recent findings reported by Kosie and Baldwin (2018b), however, provide some reassurance on this front. In their research, viewers advanced through a slideshow depicting a novel, naturalistic method of shoelace tying. The resulting dwell-time patterns revealed attentional reorganization similar to that documented in the present research. Together, these two sets of findings suggest that the attentional reorganization first documented in this research is likely generalizable to a broad range of activity sequences.

Another possible limitation of the present research arises from constraints inherent to the dwell-time paradigm itself, at least as it is currently instantiated. In principle, one could present viewers with slideshows that include as many slides as a digitized video contains (i.e., 30 frames per second, with

the present stimuli). To avoid exhausting viewers, slideshows are created by extracting slides at a regular increment (i.e., four per second), yielding relatively “low-resolution” depictions of an activity sequence. As well, slideshows are purely visual and eliminate the other sensory information (e.g., auditory, kinesthetic) found in naturalistic activity. These issues raise yet other concerns about the generalizability of our findings to action processing at large. Recent evidence has provided initial reassurance that dwell-time patterns, such as the boundary advantage, are robust to slideshow resolution (Kosie & Baldwin, 2018b). Nevertheless, it will be important to gain converging evidence for dwell-time findings via methods that enable measurement of viewers' processing of full-resolution, multimodal, streaming activity sequences.

Conclusions and future directions

In the present research, we used the dwell-time paradigm to examine how attention reorganizes as observers discover statistically defined, higher-order segmental structure within novel activity sequences. These findings open a host of additional questions. The present results are directly informative only about attentional reorganization during the processing of human activity. To what extent do similar reorganizational phenomena occur for other kinds of events? A variety of findings point to similar phenomena in language (e.g., Astheimer & Sanders, 2011) and the processing of faces (e.g., Pascalis et al., 2005). Perhaps the human mind adopts the same general strategy across domains, of reshaping attention around a discovered predictability structure, thus enhancing processing (e.g., Baldwin, 2005; Cassimatis, 2006; Christiansen & Chater, 2016). Future research might explore the extent of genuine comparability in the learning mechanisms across these domains.

Additionally, the dwell-time paradigm might be used to explore how viewers' attention is reshaped by other kinds of pattern learning, causal inferences, processing goals, and prior experience of various sorts. The dwell-time paradigm might also be used to explore how attentional reorganization develops across childhood, as well as to understand processing deficits and other individual differences that impact sociocognitive functioning in profound and ubiquitous ways. The dwell-time paradigm may also offer a better behavioral anchor than the standard forced choice discrimination task for neuroimaging methods that probe structure discovery at the neural level. These are exciting avenues for future research.

Author note The authors gratefully acknowledge Megan Angotti, Victoria McDougal, Kara Sage, Melissa Woytus, Jessica Zumbiel, and Danielle Strauss for help with data collection and coding. This research was supported by a Ruth L Kirschstein National Research Service Award

Table 1 Description of small motion elements for each of the exposure sets

SMEs	Description
Uniform Set (Single-Object/Single-Exemplar)	
blow	bottle is lifted to mouth and blown into
clean	flat hand wipes top of bottle
drink	bottle is lifted and tipped into mouth as if drinking
inspect	bottom of bottle is examined
peek	bottle is lifted to eyes and interior examined
poke	index finger is inserted and removed from top of bottle
pour	bottle is turned over as if to pour into open hand
rattle	bottle is lifted close to ear and shaken
read	finger traces over label and bottle is lifted from table as if to read
slide	bottle is pushed forward on the table and returned
touch	index finger touches side of bottle in an up-and-down motion
twirl	bottle edge is lifted from table and spun around
Variable Set (Variable-Object/Variable-Exemplar) <i>Actions also used in the Uniform-Object Set are in italics</i>	
<i>blow</i>	bottle is lifted to mouth and blown into
<i>clink</i>	bottle is tapped against side of glass
<i>drink</i>	bottle is lifted and tipped into mouth as if drinking
<i>insert</i>	bottle is put inside of glass and lifted out
<i>inspect</i>	bottom of bottle is examined
<i>peek</i>	bottle is lifted to eyes and interior examined
<i>poke</i>	index finger is inserted and removed from top of bottle
<i>pour</i>	bottle is turned over as if to pour into open hand
<i>rattle</i>	bottle is lifted close to ear and shaken
<i>scrub</i>	bottle is placed on sponge and turned
<i>stack</i>	sponge is lifted and placed on top of bottle
<i>touch</i>	index finger touches side of bottle in an up-and-down motion

(1F32HD053209-03) from the National Institute of Child Health and Human Development to the first author, as well as by a grant from the Office of Naval Research to the third author.

Appendix: Creation of the exposure and test materials

For both exposure corpora, each small motion element (SME) was filmed from the same camera angle and started and ended with the actor in the same neutral position, enabling the concatenation of SMEs in any order to result in the appearance of continuous and physically plausible intentional motion. See Table 1 for the full list of SMEs in each set. Within each set (uniform and variable), SMEs were grouped into four *actions*, with each action consisting of three randomly selected SMEs. In the uniform set, the four actions were Action 1: *pour-clean-inspect*, Action 2: *touch-blow-peek*, Action 3: *drink-twirl-read*, and Action 4: *rattle-slide-poke*. In the variable set, the

four actions were Action 1: *scrub-inspect-stack*, Action 2: *touch-poke-pour*, Action 3: *peek-clink-rattle*, and Action 4: *blow-insert-drink*.

Next, the four actions for each set were randomly ordered into a continuous sequence. Note that, across the

Table 2 Actions and part-actions used in test materials for each exposure set

Actions	Part-Actions
Single Set (Single-Object/Single-Exemplar)	
Pour–Clean–Inspect	Clean–Inspect–Touch
Touch–Blow–Peek	Blow–Peek–Pour
Drink–Twirl–Read	Read–Rattle–Slide
Rattle–Slide–Poke	Poke–Drink–Twirl
Variable Set (Variable-Object/Variable-Exemplar)	
Scrub–Inspect–Stack	Inspect–Stack–Touch
Touch–Poke–Pour	Poke–Pour–Scrub
Peek–Clink–Rattle	Rattle–Blow–Insert
Blow–Insert–Drink	Drink–Peek–Clink

exposure corpus, the ordering of SMEs within each “action” was fixed (e.g., *inspect* always followed *clean*, which always followed *pour*), meaning that the transitional probabilities from one motion element to the next within an action were 1.0, whereas the transitional probabilities as one “action” transitioned to another were, on average, .33. No action followed itself, so each of the three remaining possible action segments had an equal probability of following an initial action segment. Thus, low transitional probabilities served as a clue to the boundaries between “action” segments. To ensure that the starting and ending points of the videos could not serve as a clue to the statistical structure of the action stream, the exposure corpus always started and ended in the middle of an action. For example, the uniform set began with Action 1, *pour-clean-inspect*, and the first SME shown was *clean*. The uniform set ended with Action 1, and the last SME shown was *pour*.

The actions in the uniform set took less time than those in the variable set, so to ensure exposure corpora of equivalent length, the corpora contained different numbers of tokens of each action. The uniform exposure set contained 40 tokens of each action, yielding a 160-action sequence (480 SMEs) and a total time of 10 min, 20 s. The variable set contained 28 tokens of each action, yielding a 112-action sequence (336 SMEs) and a total time of 10 min, 1 s. For the variable set, each token represented a unique version of that action. Because there were 28 exemplars of each SME in the variable set, each token of a given action was constructed by randomly selecting one of the 28 variants of the SMEs composing that action. Thus, the first token of Action 1, *scrub-inspect-stack*, might include Variant 1 of *scrub*, Variant 14 of *inspect*, and Variant 19 of *stack*. But the second token of the same action would include different variants of each of the component SMEs. To enhance the continuity of the motion stream, the transitions between individual SMEs (both within and across actions) were smoothed using the Overlap transition in iMovie (version 5.0.2).

A discrimination test following the learning-phase tested whether observers recognized “actions” as being more familiar than foils, which consisted of “part-actions.” Possible three-element sequences that spanned action boundaries, and thus featured transitional probabilities of <1.0. Video clips of each of the four actions used in the exposure corpus were extracted for use in the test phase. We also created part-actions sequences of three SMEs that spanned action boundaries (see Table 2 for a list of the actions and part-actions). The 16 test trials were created by exhaustively pairing each action with each part-action. These pairings were quasirandomly ordered: They were random, with the stipulation that no test pair could contain an action or part-action that a prior test pair had contained.

References

- Asstheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, 49, 3512–3516.
- Baldwin, D. (2005). Discerning intentions: Characterizing the cognitive system at play. In B. Homer & C. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 117–144). Mahwah, NJ: Erlbaum.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382–1407.
- Baldwin, D. A., & Baird, J. A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5, 171–178.
- Baldwin, D., & Pederson, E. (2016). *Attentional enhancement at event boundaries*. Poster presented at the meeting of the Cognitive Science Society, Philadelphia, PA.
- Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, 83, 62–78. <https://doi.org/10.1016/j.jml.2015.04.004>
- Cassimatis, N. L. (2006). A cognitive substrate for achieving human-level intelligence. *AI Magazine*, 27, 45–56.
- Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, e62:1–52. <https://doi.org/10.1017/S0140525X1500031X>
- Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236–243. [https://doi.org/10.1016/S1364-6613\(00\)01662-4](https://doi.org/10.1016/S1364-6613(00)01662-4)
- Hard, B. M., Recchia, G., & Tversky, B. (2011). The shape of action. *Journal of Experimental Psychology: General*, 140, 586–604.
- Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, 34, 1221–1235.
- Hartley, R. V. L. (1928). Transmission of information. *Bell System Technical Journal*, 17, 535–563.
- Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, 20, 848–864.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, 7, e36399. <https://doi.org/10.1371/journal.pone.0036399>
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, 85, 1795–1804. <https://doi.org/10.1111/cdev.12263>
- Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 461, 145–149.
- Koo, H. (2008). Change in the adult phonological processing system by learning non-adjacent phonotactic constraints from brief experience: An experimental and computational study. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 69, 591.
- Kosie, J., & Baldwin, D. (2018a). Attention rapidly reorganizes to structure in a novel activity sequence. *Cognition*. Accepted manuscript pending minor revision.
- Kosie, J., & Baldwin, D. (2018b). Attentional profiles linked to event segmentation are robust to missing information. *Cognitive Research: Principles & Implications* (Special issue on attention in natural and mediated realities). Invited resubmission.
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12, 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>

- Mayberry, C. R., Livesey, E. J., & Dux, P. E. (2010). Rapid learning of rapid temporal contexts. *Psychonomic Bulletin & Review*, 17, 417–420. <https://doi.org/10.3758/PBR.17.3.417>
- Meyer, M., & Baldwin, D. A. (2011). Statistical learning of action: The role of conditional probability. *Learning and Behavior*, 39, 383–398.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 28, 28–38.
- Newton, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 12, 436–450. [https://doi.org/10.1016/0022-1031\(76\)90076-7](https://doi.org/10.1016/0022-1031(76)90076-7)
- Newton, D., Engquist, G. A., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology*, 35, 847–862. <https://doi.org/10.1037/0022-3514.35.12.847>
- Olson, I. R., & Chun, M. M. (2001). Temporal contextual cuing of visual attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1299–1313. <https://doi.org/10.1037/0278-7393.27.5.1299>
- Pani, J. R. (2000). Cognitive description and change blindness. *Visual Cognition*, 7, 107–126.
- Pascalis, O., Scott, L. S., Kelly, D. J., Shannon, R. W., Nicholson, E., Coleman, M., & Nelson, C. A. (2005). Plasticity of face processing in infancy. *Proceedings of the National Academy of Sciences*, 102, 5297–5300.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238.
- Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B*, 364, 3711–3735.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, 12, 110–114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month old infants. *Science*, 274, 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621. <https://doi.org/10.1006/jmla.1996.0032>
- Sanders, L. D., Amoral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia*, 47, 1183–1186.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, 85, 1821–1826. <https://doi.org/10.1111/cdev.12247>
- Swallow, K. M., & Zacks, J. M., (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review*, 15, 116–122. <https://doi.org/10.3758/PBR.15.1.116>
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138, 236–257. <https://doi.org/10.1037/a0015631>
- Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134, 552–564. <https://doi.org/10.1037/0096-3445.134.4.552>
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21, 1934–1945.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28, 979–1008.
- Zacks, J. M., Kumar, S., Abrams, R. A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, 112, 201–216.
- 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. <https://doi.org/10.1037/0033-2909.133.2.273>
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, 130, 29–58. <https://doi.org/10.1037/0096-3445.130.1.29>