Infants Segment Continuous Events Using Transitional Probabilities

Aimee E. Stahl,
Johns Hopkins University

Alexa R. Romberg,
Indiana University

Sarah Roseberry,
University of Washington

Roberta Michnick Golinkoff, and
University of Delaware

Kathryn Hirsh-Pasek
Temple University

Abstract

Throughout their first year, infants adeptly detect statistical structure in their environment. However, little is known about whether statistical learning is a primary mechanism for event segmentation. This study directly tests whether statistical learning alone is sufficient to segment continuous events. Twenty-eight 7- to 9-month-old infants viewed a sequence of continuous actions performed by a novel agent in which there were no transitional movements that could have constrained the possible upcoming actions. At test, infants distinguished statistically intact units from less predictable ones. The ability to segment events using statistical structure may help infants discover other cues to event boundaries, such as intentions, and carve up the world of continuous motion in meaningful ways.

Imagine a frequent event in the life of an infant, like bath time. A parent might place the infant in the tub, open a bottle and put soap in their hands, wash the infant, and rinse. These actions are likely to occur in the same order each time this event happens. On a daily basis, infants observe and engage in routines like this that are comprised of reliable sequences of actions. The event of bath time may be followed by other events that also have predictable structure, like putting on pajamas or book reading. How do infants know when one event ends and another begins? Segmenting events into units is critical for many skills, including anticipating future actions, imitating others, categorizing events, and learning words that label those actions. An outstanding question for developmental scientists is how infants parse the action sequences that make up events in a way that scaffolds these skills.

Correspondence concerning this article should be addressed to Aimee E. Stahl, Department of Psychological & Brain Sciences, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. Electronic mail may be sent to astahl4@jhu.edu.

One hypothesis for how event perception progresses is that infants begin with basic, domain-general learning mechanisms that allow them to group actions based on the sequential predictability of the actions they observe (Baldwin & Baird, 2001; Baldwin, Baird, Saylor, & Clark, 2001; Roseberry, Richie, Hirsh-Pasek, Golinkoff, & Shipley, 2011). Infants could use these initial groupings to discover more abstract cues to event structure, such as the actor's intentions, that are known to play a role in adults' global event segmentation (e.g., Wilder, 1978; Zacks & Tversky, 2001; Zacks, 2004). A similar arc has been proposed for word segmentation. Infants initially use basic perceptual learning to parse the speech stream, in which they perceive highly predictable sequences of syllables as more word-like than less predictable sequences (e.g., Aslin, Saffran, & Newport, 1998; Graf Estes, Evans, Alibali, & Saffran, 2007; Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996). Infants then generalize across familiar words to find more language-specific cues to word boundaries, such as lexical stress (Sahni, Seidenberg, & Saffran, 2010; Thiessen, Kronstein & Hufnagle, 2013; Thiessen & Saffran, 2003).

Several studies have demonstrated infants' visual sequence learning skills by testing whether they can track regularities in static features like shape and color (Bulf, Johnson, & Valenza, 2011; Kirkham, Slemmer, & Johnson, 2002). However, event sequences consist of dynamically changing movements rather than static features. Eight-month-old infants are sensitive to the sequential statistics of actions performed by a human agent (Roseberry et al., 2011). Infants viewed a sequence of hand motions in which some motions reliably followed other motions, forming units that were combined into larger sequences. Similar to real-world actions, unique transitional movements connected each hand motion to the one that came after. For example, transitioning from the motion of pressing palms together to the motion of forming an "X" with one's arms requires that one hand pass in front of the other. This motion is markedly different than transitioning from pressing palms to stacking one's fists. These unique transitions provided cues to the sequential structure much like co-articulation in fluent speech, where the production of one sound is influenced by the pronunciation of the preceding or following sound. In Roseberry et al. (2011), the next action was constrained to those physically compatible with the trajectory of the transitional motion from the preceding action. Thus, infants had two cues to the sequential structure: Sequential regularities and the transitional movement that connected one motion to the next.

There is no doubt that such physical constraints are present in everyday actions and likely provide useful information about upcoming motions. However, unique transitions are not required for auditory sequence learning with either words (e.g., Gómez, 2002; Lany & Gómez, 2008) or tones (e.g., Saffran, Johnson, Aslin & Newport, 1999). Therefore, the current study sought to test whether infants can segment action sequences based solely on their sequential predictability without the additional cue of transitional movements that physically constrain the upcoming motion.

A finding that infants can segment actions based only on sequential predictability would bolster the claim that statistical learning operates similarly across domains. Conversely, a finding that infants must rely on physical constraints as well as sequential structure would raise the possibility that statistical learning of action sequences relies on a different mechanism than statistical learning in other domains. Although researchers have concluded

that finding predictable structure in sequential stimuli is a fundamental learning mechanism (Kirkham, Slemmer & Johnson, 2002; Krogh, Vlach & Johnson, 2013), it has yet to be established that the same learning mechanism accounts for all types of statistical learning (Conway & Christiansen, 2006; Thiessen, Kronstein & Hufnagle, 2013).

One recent study found that *adults* segmented a continuous action sequence based on sequential predictability alone (Baldwin, Andersson, Saffran, & Meyer, 2008). The materials in that study consisted of videos of an actor performing simple actions. The actor started and finished each action in the same position, thereby eliminating additional cues about the upcoming action. The current study applies this design technique to ask whether infants also can segment actions based solely on sequential predictability. Additionally, we tested whether infants can segment actions performed by a completely novel (non-human) agent or whether segmentation relies on the familiarity of the stimulus (e.g., hands, such as in Roseberry et al., 2011).

In the present study, infants were shown an animated agent (a starfish with eyes) performing whole-body actions (e.g., spin, twist, bend). Each action started and ended in a fully extended "star" position so that actions could flow naturally and continuously without transitional movements providing additional cues to the sequential structure of the units. Actions were arranged into four 3-action units that formed the full, continuous sequence. If infants' ability to track sequential structure in actions relies on physically constraining transitional motions between actions, or on stimulus familiarity, they should fail to segment the sequence and therefore not recognize the 3-action units at test. If infants possess robust perceptual learning skills for action segmentation, as they do for language, infants should discriminate between the highly predictable 3-action units and sequences that are less predictable.

Method

Participants

Participants were 28 monolingual, English-reared infants (13 female) between the ages of 7 and 9 months (M = 236 days, SD = 18.4, range = 209-266 days). One additional infant was excluded for being distracted by external noise. Familiarization data from four of the 28 infants were not saved due to a computer error. The final sample was predominantly Caucasian and middle-class.

Stimuli and Procedure

Infants sat on a parent's lap facing the monitor displaying the stimulus. Parents were instructed to close their eyes and refrain from speaking. An observer, blind to the stimulus sequence viewed by the infant, watched the infant on another monitor and indicated by key press when the infant was looking at the screen. Two computers controlled the stimuli through Habit X software (Cohen, Atkinson, & Chaput, 2004): One controlled the auditory attention-getter during familiarization; the other controlled visual stimuli during familiarization and audio and visual presentations during the test phase.

Familiarization—Stimuli consisted of a purple animated starfish with eyes in the center of a black screen. The starfish performed 12 actions (bow, clap, back-flip, turn, jumping-jack, back-bend, side-bend, twist, arm circles, split, toe-touch, scrunch). Each action was a 1-second cycle that began and ended in a fully extended "star" position. The return to the "star" position allowed actions to be combined in any order and ensured that transitional movements between actions did not provide cues to the sequence structure.

The 12 actions were organized into four unique 3-action units (Figure 1). Within-unit transitional probabilities (TPs) were 1.0. The event corpus consisted of a total of 42 units. Two of the 3-action units occurred 14 times (high frequency units) and the other two occurred 7 times (low frequency units). The units were arranged in the sequence to create "part-units" that spanned the last action of one high frequency unit and the first two actions of the other high frequency unit. These part-units had an internal TP of 0.5 between the first and second action, and occurred the same number of times as the low frequency action units (Aslin, Saffran & Newport, 1998). Infants viewed one of two familiarization corpora counterbalanced for which action units were high and low frequency.

The familiarization video played continuously for its 126-second duration. If infants looked away from the screen for 1 s, an auditory attention-getter (a futuristic noise) played until they resumed looking at the screen. The video continued to play during the auditory attention-getter.

Test—Immediately after familiarization, infants viewed a series of 12 test trials, presented in three blocks of four test sequences. Two of these sequences were the low frequency units and the other two were the part-units described above. Because the low frequency units and the part-units occurred with equal frequency in the familiarization corpus, successful discrimination required that infants recognize the differing TP structure between these items. The low frequency intact units in one corpus were the part-units in the other, and vice versa, which ensured that infants did not simply prefer particular action sequences over others during the test trials.

The test trials were presented in one of four preset orders that varied which test item was presented first; the same type of trial never appeared more than twice in a row. Each test trial was preceded by an attention-getter (a looming geometric pattern accompanied by a futuristic sound). The test trial began once the infant looked at the screen. Test sequences were looped until infants looked away for 1 s or for a maximum duration of 15 s.

Results

Infants were generally attentive during the experiment. The average looking time to the familiarization stimulus was 94.9 s (SD=17.2 s, range = 57.1 to 119.8 s) of the 126-second video. Because each individual action was one second long, the minimum time required to distinguish the test items was 1.5 s, which was enough to see the first action and the beginning of the second (the crucial pair to distinguish intact from part-units). Test trials on which infants looked for less than 1.5 s were excluded. This led to the exclusion of 12 trials (3.5%). All infants contributed at least 4 test trials of each trial type to the analysis.

A preliminary analysis revealed no significant differences in looking times for the different counterbalanced corpora ($p > .18$) and so all analyses were collapsed across corpora. If infants used TPs to segment the 3-action units from the continuous stream, then part-units with a mid-sequence dip in TP should be novel compared to intact units. As predicted, infants looked longer to part-units (M=6.70 s, SD=2.68 s) than to intact units (M=5.91 s, SD=2.75 s), and this difference was statistically significant, t(27)=2.09, $p = .046$. Group means are shown in Figure 2; 19 out of 28 infants (68%) looked longer to the part-units than to the intact units (two-sided binomial test, $p = .087$).

Because infants were free to look away from the screen during the familiarization phase, they varied in how much time they spent viewing the action sequence. Familiarization time was positively correlated with *total* looking time during test, indicating that some infants were more attentive overall, r(22)=0.45, $p = .03$. However, the amount of time infants looked during familiarization was not correlated with their looking time difference to intact and part-units, r(22) = 0.08, $p = .68$. Thus, it was not the case that infants who looked more during familiarization showed stronger novelty preferences at test. Infants' age and gender were also unrelated to their difference scores (p's > .6).

Discussion

This study sought to determine whether infants can use transitional probabilities alone to segment events, and our results demonstrate infants' keen sensitivity to statistical structure in action sequences. Infants successfully segmented a continuous stream of actions performed by a novel agent, looking significantly longer to part-units than to intact units. Importantly, the part-units and intact units used to test segmentation occurred with equal frequency during familiarization. Thus, infants must have utilized the probability that one action followed another to discriminate between intact and part-units.

Several studies have demonstrated infants' sensitivity to the distribution of static features in visual sequences (e.g., Kirkham et al., 2002; Fiser & Aslin, 2002). The current study crucially extends this body of work by revealing that infants track sequential regularities in continuously changing actions without predictive transitional movements between them (see also Roseberry et al., 2011). Our results provide additional evidence that statistical learning is a fundamental learning mechanism that operates over many domains in similar ways, and that tracking distributional information may play an important role in event segmentation in infancy. However, given that not all infants looked longer to part-units than intact units, it is likely that there are individual differences in infants' statistical learning abilities that contribute to their event segmentation.

These findings add to a growing body of literature investigating the skills underlying infant event segmentation and suggest that perceptual learning may provide the foundation for learning higher-level cues for action processing, such as the agent's intention (e.g., Baldwin et al., 2001; Saylor, Baldwin, Baird, & LaBounty, 2007). Before infants have top-down knowledge of intentions, they may begin to carve up events in the world based on sequential predictability. The observation of reliable sequences of actions has been implicated in sequence learning by non-human species. After repeated viewings of action sequences,

young gorillas can parse individual actions and extract their sequential predictability without necessarily understanding the intention of the action sequence. This "string parsing" may ultimately help them to produce the sequence of behaviors themselves (Byrne, 1999). Infants in our experiment faced a similar challenge, as there were no cues to causes, goals, or intentions. Rather, they had to extract the statistical regularities only by observing the sequence of actions.

The ability to segment a continuous stream of actions enables infants to perceive events as consisting of coherent units. This may make processing more efficient by allowing them to predict downstream events and form hierarchical event representations (e.g., first comes bath time, then comes book reading, both of which are part of the larger event of "getting ready for bed"). Event segmentation may even enhance memory for those events (Zacks, Speer, Vettel, & Jacoby, 2006; Zacks & Swallow, 2007). Furthermore, in order for infants to themselves *produce* action sequences (e.g., bathing themselves or a baby doll), they need to recognize and parse the individual components (e.g., wash then rinse). Tracking sequential predictability might be one means by which they achieve comprehension and then production of meaningful action sequences.

Several open questions remain to be addressed in future research. First, how does sequential predictability influence infants' ability to learn a label for a novel action sequence? Infants' formation of novel label-object mappings is facilitated when labels consist of predictable syllable sequences based on immediately prior linguistic exposure (Graf Estes et al., 2007; Hay, Pelucchi, Graf Estes, & Saffran, 2011). Likewise, infants may more readily map labels to novel events with higher internal consistency based on their prior experience. Second, future studies should examine whether and when infants generalize links created by distributional information from one agent to another. Once infants map labels to such sequences, they will need to extend them to new contexts in order to use language generatively (e.g., the action sequence of bath time can be labeled as such whether dad or the babysitter performs it). Finally, how does the presence of other cues, like intentions (Baldwin et al., 2001), familiarity (Hespos, Saylor & Grossman, 2009), or goals (Loucks & Meltzoff, 2013), influence the role of sequential predictability in event segmentation? In natural language processing, word segmentation is facilitated by the conjunction of multiple cues, such as sequential predictability, familiar words, and isolated utterances (Mersad & Nazzi, 2012; Lew-Williams, Pelucchi, & Saffran, 2011). Tracking distributional information in action sequences likely contributes to event processing, but event segmentation will no doubt be most robust when multiple cues align.

The present work suggests that during the first year of life infants possess the ability to segment events using sequential predictability. This domain-general mechanism may support the development of domain-specific and higher-order processes, a vital precursor to carving up the world in meaningful ways.

Acknowledgments

References

Aslin RN, Saffran JR, Newport EL. Computation of conditional probability statistics by 8-month-old infants. Psychological Science. 1998; 9:321–324.

Baldwin D, Andersson A, Saffran JR, Meyer M. Segmenting dynamic human action via statistical structure. Cognition. 2008; 106:1382–1407. [PubMed: 18035346]

Baldwin DA, Baird JA. Discerning intentions in dynamic human action. Trends in Cognitive Sciences. 2001; 5:171–178. [PubMed: 11287271]

Baldwin DA, Baird JA, Saylor MM, Clark MA. Infants parse dynamic action. Child Development. 2001; 72:708–717. [PubMed: 11405577]

Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. Cognition. 2011; 121:127–132. [PubMed: 21745660]

Byrne RW. Imitation without intentionality. Using string parsing to copy the organization of behaviour. Animal Cognition. 1999; 2:63–72.

Cohen, LB.; Atkinson, DJ.; Chaput, HH. Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas; 2004.

Conway CM, Christiansen MH. Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. Psychological Science. 2006; 17:905–912. [PubMed: 17100792]

Fiser J, Aslin RN. Statistical learning of new visual feature combinations by infants. Proceedings of the National Academy of Sciences. 2002; 99:15822–15826.

Gómez RL. Variability and detection of invariant structure. Psychological Science. 2002; 13:431–436. [PubMed: 12219809]

Graf Estes KM, Evans J, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science. 2007; 18:254–260. [PubMed: 17444923]

Hay JF, Pelucchi B, Graf Estes K, Saffran JR. Linking sounds to meaning: Infant statistical learning in a natural language. Cognitive Psychology. 2011; 63:93–106. [PubMed: 21762650]

Hespos SJ, Saylor M, Grossman S. Infants' ability to parse continuous actions. Developmental Psychology. 2009; 45:575–585. [PubMed: 19271840]

Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: Evidence of a domain general learning mechanism. Cognition. 2002; 83:B35–B42. [PubMed: 11869728]

Krogh L, Vlach HA, Johnson SP. Statistical learning across development: Flexible yet constrained. Frontiers in Language Sciences. 2013; 3:598.

Lany J, Gomez RL. Twelve-month-olds benefit from prior experience in statistical learning. Psychological Science. 2008; 19:1274–1252.

Lew-Williams C, Pelucchi B, Saffran JR. Isolated words enhance statistical language learning in infancy. Developmental Science. 2011; 14:1323–1329. [PubMed: 22010892]

Loucks J, Meltzoff AN. Goals influence memory and imitation for dynamic human action in 36-month-old children. Scandinavian Journal of Psychology. 2013; 54:41–50. [PubMed: 23121600]

Mersad K, Nazzi T. When Mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. Language Learning & Development. 2012; 8:303–315.

Romberg AR, Saffran JR. Statistical learning and language acquisition Wiley Interdisciplinary Reviews. Cognitive Science. 2010; 1:6:906–914. [PubMed: 21666883]

Roseberry S, Richie R, Hirsh-Pasek K, Golinkoff RM, Shipley T. Babies catch a break: Seven- to 9-month-olds track statistical probabilities in continuous dynamic events. Psychological Science. 2011; 22:1422–1424. [PubMed: 22020978]

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science. 1996; 274:1926–1928. [PubMed: 8943209]

Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999; 70:27–52. [PubMed: 10193055]

Sahni SD, Seidenberg M, Saffran JR. Connecting cues: Overlapping regularities support cue discovery in infancy. Child Development. 2010; 81:727–736. [PubMed: 20573101]

Saylor M, Baldwin D, Baird J, LaBounty J. Infants' on-line segmentation of dynamic human action. Journal of Cognition and Development. 2007; 8:113–128.

Thiessen ED, Kronstein AT, Hufnagle DG. The extraction and integration framework: A two-process account of statistical learning. Psychological Bulletin. 2013; 139:792–814. [PubMed: 23231530]

Thiessen ED, Saffran JR. When cues collide: Statistical and stress cues in infant word segmentation. Developmental Psychology. 2003; 39:706–716. [PubMed: 12859124]

Wilder DA. Effect or predictability on units of perception and attribution. Personality and Social Psychology Bulletin. 1978; 4:281–284.

Zacks JM. Using movement and intentions to understand simple events. Cognitive Science. 2004; 28:979–1008.

Zacks JM, Speer NK, Vettel JM, Jacoby LL. Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychology & Aging. 2006; 21:466–482. [PubMed: 16953710]

Zacks JM, Swallow KM. Event segmentation. Current Directions in Psychological Science. 2007; 16:80–84. [PubMed: 22468032]

Zacks JM, Tversky B. Event structure in perception and conception. Psychological Bulletin. 2001; 127:3–21. [PubMed: 11271755]
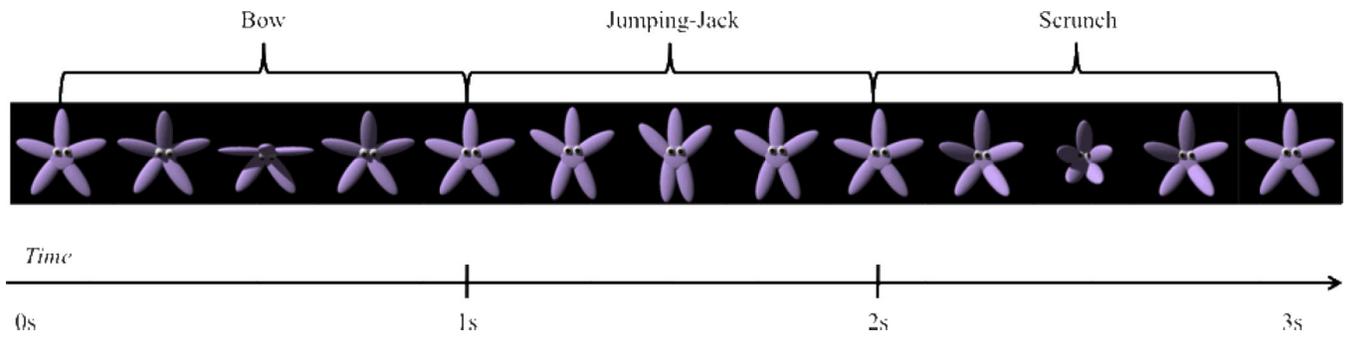
Figure 1.
Sample sequence of actions from the familiarization corpus. These static images represent what was shown in the dynamic events.
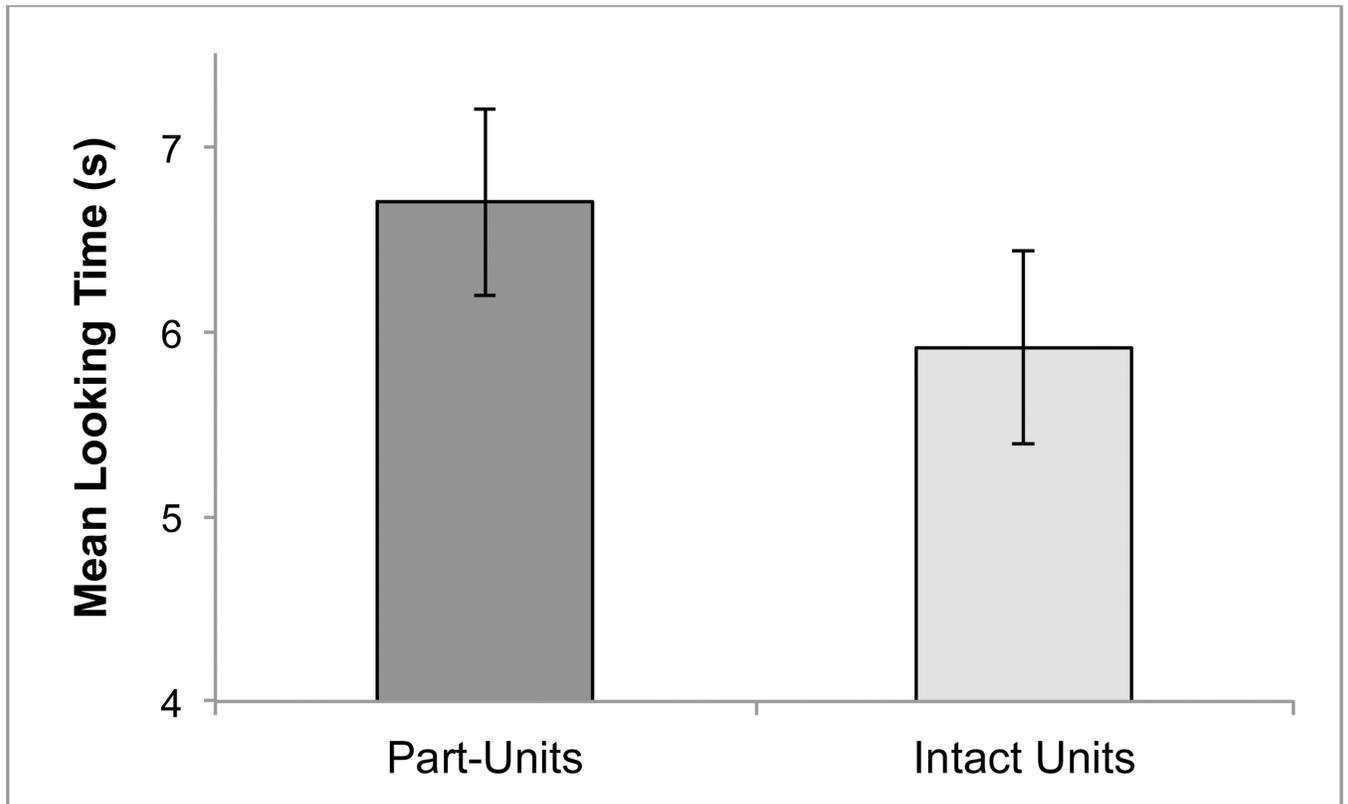
Figure 2.
Mean looking times at test for part-units and intact units. Bars represent standard error of the mean.
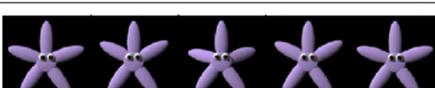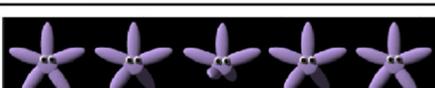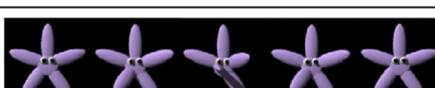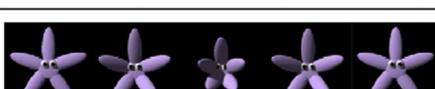
Appendix.
Static Depictions of Each Dynamic 1-second Action in the Familiarization Corpus