Abstract

We present two habituation experiments that examined 20- and 26-month-olds’ ability to engage in second-order correlation learning for static and dynamic features whereby learned associations between two pairs of features (e.g., P and Q, P and R) are generalized to the features that were not presented together (e.g., Q and R). We also present results from an associative learning mechanism that was implemented as an autoencoder parallel distributed processing (PDP) network in which second-order correlation learning is shown to be an emergent property of the dynamics of the network. The experiments and simulation demonstrate that 20- and 26-month-olds as well as neural networks are capable of second-order correlation learning in a category context for internal features of dynamic objects. However, the model predicts—and Experiment 3 demonstrates—that 20- and 26-month-olds are unable to encode second-order correlations in a non-category context for dynamic objects with internal features. It is proposed that the ability to learn second-order correlations represents a powerful but as yet unexplored process for generalization in the first years of life.

**Introduction**

A key debate in the developmental literature concerns how infants learn to represent animates (i.e., people, animals, and insects) and inanimates (e.g., vehicles, furniture, plants, and tools) and how this knowledge generalizes to the varied and complex things they encounter in the world. There is now considerable evidence that infants’ earliest representations for animals, vehicles, plants, tools, and people are grounded in static and dynamic perceptual, or surface, features (see e.g., Quinn & Eimas, 1997; Younger & Cohen, 1986). For example, infants as young as 3 months of age within the familiarization paradigm can form categorical representations of pictures for dogs that exclude cats (Quinn, Eimas, & Rosenkrantz, 1993) and for pictures of mammals that exclude birds, fish, and furniture (Behl-Chadha, 1996). Moreover, infants at 14 and 18 months of age form categories of animals and vehicles in the sequential touching paradigm on the basis of parts such as legs and wheels (Rakison & Butterworth, 1998a). That the categories infants form are susceptible to subtle changes in the features of the stimuli suggests that these representations are based on bottom-up (i.e., perceptually driven) processes rather than top-down (i.e., conceptually driven) ones (French, Mareschal, Mermillod, & Quinn, 2004; Rakison & Butterworth, 1998a, 1998b).

Infants are also highly sensitive to the statistical regularities to which they are exposed across a range of perceptual inputs (e.g., Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). As a result, it is generally assumed that infants’ initial representations for things in the world are formed on the basis of correlations—or bundles—of static and dynamic features. These bundles of features—for example, beaks and feathers or hands and feet—tend to appear together in the environment and are encoded by domain-general associative learning mechanisms. A large body of evidence suggests that from birth infants encode correlations among static features embedded in both artificial and naturalistic stimuli (e.g., Fiser & Aslin, 2001; Gogate & Bahrick, 1998; Madole & Cohen, 1995; Slater, Mattock, Brown, & Bremner, 1991). Slater et al. (1991), for example, found that newborns who were familiarized to two stimulus compounds—namely, a green vertical stripe and a red diagonal stripe—learned the relation between the two features (color and angle of the stripe) rather than each feature independently (e.g., just the color or the angle of the stripe). Moreover, Younger and colleagues (Younger & Cohen, 1986; Younger & Gotlieb, 1988), for example, have shown across a variety of experimental paradigms that 10-month-old infants are sensitive to correlated features in a category context for artificial and realistic animal stimuli as well as realistic color photographs of animals. Thus, the initial representations for objects and entities in the world may be grounded in correlations among the static and dynamic features of objects, which suggests that associative learning may be a primary mechanism that underpins early knowledge acquisition for animates and inanimates.

One issue that remains largely untested is how and when infants and young children learn about correlations between features that are rarely, if ever, observed together; that is, there is currently little evidence that attests to whether infants and children can learn correlations between features that are only indirectly correlated. This ability—labeled *second-order correlation learning* (Yermolayeva & Rakison, 2016)—is important because children must make inferences about the presence (or absence) of features from intermittently available stimuli. For example, if children learn that things with hands are goal-directed and that things with hands have eyes, can they infer on the basis of associations between these features that things with eyes move in a goal-directed manner? This kind of second-order correlation learning is likely applied by infants in a number of contexts including causal learning (Walker & Gopnik, 2014), base-rate learning (Dewar & Xu, 2010), transitive learning (Mou, Province, & Luo, 2014), and language learning (Sandoval & Gomez, 2013), among other things. The aim of the two experiments and simulation presented here was to establish whether infants and young children can learn such correlations for dynamic, moving stimuli with internal features that mirror in an abstract sense the movement of animates and inanimates in the real world.

Despite the relative dearth of research on this issue, there are at least three reasons to believe that infants are capable of second-order correlation learning. First, it is not necessary that two cues have temporal contiguity for associative learning to occur; such learning has been demonstrated in *trace conditioning* studies in which the conditioned stimulus is not physically present when the unconditioned stimulus is presented (Dwyer, Mackintosh, & Boakes, 1998; [Pavlov, 1927)](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2288639/#R20). For example, in one study by Cuevas, Rovee-Collier, and Learmonth (2006; see also Barr, Marrot, & Rovee-Collier, 2003), 6-month-old infants were taught that (1) two hand puppets (A and B) went together; (2) that a mobile could be moved by the infant’s kicking that went with one of two particular cribs; and (3) that one of the hand puppets (A) went with one of the cribs. Based on these three available correlations, Cuevas et al. demonstrated that infants associated the second hand puppet (B) with the mobile—even though they did not experience the two together—through the activated memories of puppet A and the crib context. Second, infants are able to learn nonadjacent dependencies among syllables in which two related syllables have an intervening syllable that is unrelated. Thus, 6-month-olds can track nonadjacent dependencies among vowels in natural language, and 10-month-olds can track nonadjacent relationships among consonants (Gonzalez-Gomez & Nazzi, 2012; for a review see Sandoval & Gomez, 2013).

Third, there exists one series of studies that has examined infants’ ability to learn second-order correlations. Yermolayeva and Rakison (2016) used the generalized imitation procedure (see Mandler & McDonough, 1996; Rakison, 2005b) to examine whether 7-, 9-, and 11-month-olds can learn such associations for static 3D objects. In the task, infants were presented serially with two sets of two 3D objects, where each set had the same body (i.e., a green cube or a blue cone) and each object in each trial was attached to a different external part (e.g., the green cube had a red cross on its top in one trial and a blue handle shape on its side in another trial). At test, subjects were given an object with a novel body (e.g., a red triangle) that possessed the two external parts that were previously attached to the same body (e.g., the red cross and the blue handle)—the *consistent* test item—and an object with the same novel body that possessed one external part from one set (e.g., the red cross) and one from another (e.g., a green knob; the *inconsistent* test item). Note that a preference for either test item is evidence of second-order correlation learning because the features of each object were presented equally during familiarization. Thus, longer looking to either object could only occur if infants learned the second-order correlation among the parts of the objects. Across four experiments it was found that 7-month-olds examined the inconsistent test item longer than the consistent test item, that 9-month-olds showed no preference for either test item, and 11-month-olds examined the consistent test item longer than the inconsistent one. This finding suggests that infants as young as 7 months of age are capable of second-order correlation learning for external static parts of static objects within the generalized imitation paradigm. It also demonstrated a novel shift in the pattern of interest whereby younger infants examined the inconsistent item longer than the consistent item and the older infants demonstrated the opposite pattern of examining.

In conjunction, these studies illustrate that infants are able to learn that multiple features are associated with a static object even if those features are not presented simultaneously. Thus, by 7 months of age infants can infer an association between features R and Q based on the correlation between P and Q and between P and R. It remains to be seen, however, if and when infants can apply such learning to the kinds of dynamic stimuli (e.g., people, animals, vehicles)—whose conceptual roles are defined by their category membership—that they observe in the everyday world. If infants and children use second-order correlation learning to generalize their knowledge to new instances, this process would have to operate not only on individual exemplars (Yermolayeva & Rakison, 2016) but also across category members. With this in mind, the goals of the current experiments and computational model were twofold. First, they were designed to test whether infants and young children can engage in the kind of second-order correlation learning described above for dynamic, moving stimuli and can do so in a category context. Second, they were designed to examine the developmental timetable for the emergence of this ability for moving, dynamic stimuli where the features and movement of the objects may not be observed together. Third, they were designed—through the implementation of a computational model—to explain the pattern of behavior observed in Yermolayeva and Rakison (2016) and in the current work.

In Experiment 1, participants at 20 and 26 months of age were habituated to six events that included category information about two separate correlations between features (see Figure 1). Two events (labeled the static object events) displayed the relation between two stationary objects’ shape and their surface feature (e.g., a blue square has a yellow heart and a red circle has a white cross), and the other four events (labeled the dynamic object events) displayed the dynamic relation between an object’s color and shape and its movement path (e.g., blue objects [a circle and a pentagon] move curvilinearly; red objects [a square and a chair-shape] move rectilinearly). In the test phase, participants were shown events with a novel moving objects (a pink flowerpot shape) that included surface features that were either inconsistent (e.g., an object with a white cross moves curvilinearly) or consistent (e.g., an object with a white cross moves linearly) with the second-order correlation that could have been extracted from the habituation events. Experiment 2 presents a computational model that was implemented to explain the results of Experiment 1 and to make predictions about infants’ behavior in a non-category context version of the task. Experiment 3 was designed to test the predictions of the computational model with 20- and 26-month-olds in a task in which subjects were given second-order correlation information about only two moving objects—in a non-category context—rather than four.

Note that in contrast to many studies with the habituation procedure, it is indeterminate whether participants should show a preference for the consistent stimulus relative the inconsistent one or vice versa if they engaged in second order correlation learning. Recall that in the studies by Yermolayeva and Rakison (2016), younger infants showed a preference for the consistent test item over the inconsistent one and older infants showed the opposite direction of preference. On the one hand, subjects in the current experiments may look longer at the consistent event than the inconsistent event to help complete their encoding to memory of the previously unseen correlation (Hunter & Ames, 1988; see also Kidd, Piantadosi, & Aslin, 2012). On the other hand, participants may look longer at the inconsistent event than the consistent event because they learned the unseen correlation and the former violates that correlation whereas the latter does not. Regardless, a preference in either direction must mean that participants encoded the unobserved correlation to some degree during the habituation phase of the experiment.

**Experiment 1**

The goal of this experiment was to determine whether 20- and 26-month-olds can infer a correlation among static and dynamic features that they do not observe but that is implied by two other, overlapping correlations. These ages were tested because, as noted above, it is between 16 and 24 months of age that infants generalize a learned association to other correlated features (Rakison, 2005a,b; Rakison, 2006). Infants were presented with category information for the dynamic aspects of each object—two objects for each movement—for two reasons: first, because earlier work examined second-order correlation learning only for individual objects (Yermolayeva & Rakison, 2016), and second because if infants apply second-order correlation learning to the movement of animates and inanimates then it must be used to extract commonalities across category members. Note, however, that category learning was not tested in the current experiment; rather, the goal of the experiment was to determine whether infants and young children can engage in second-order correlation learning when multiple instances with shared properties—in this case, the color of the body of the moving object—are presented.

**Method**

**Participants.** The participants were twenty four healthy full-term 20-month-olds (mean age 19 months 29 days; range = 19;15 to 20;26) and twenty four healthy full-term 26-month-olds (mean age 26 months 9 days; range = 25;14 to 26:28). There were an equal number of males and females in the 20-month-old group and 13 females and 11 males in the 26-month-old group. The majority of participants were White and of middle socioeconomic status. Data from an additional 36 participants were excluded from the final sample, 12 because of fussing, 10 for a failure to habituate, 4 because of technical problems, 6 because of an experimenter error, and 4 for looking times that were 2 SD from the condition mean. Although the dropout rate from fussiness and a failure to habituate was relatively high, it is not atypical of habituation studies with older infants and young children (e.g., Rakison & Poulin-Dubois, 2002) because such children can become easily bored or agitated in a habituation context. Participants were recruited through birth lists obtained from a private company and were given a small gift for their participation.

We tested 24 infants in each age group for two reasons. First, a statistical power analysis was performed for sample size estimation, based on data from Yermolayeva and Rakison (2016). With an alpha = .05 and power = 0.80, the projected sample size needed with the effect size in Yermolayeva and Rakison (2016) is approximately N = 20 for within group comparison. Second, 24 subjects were used because this allowed for complete counterbalancing of the presentation of the stimuli across infant subjects.

**Stimuli and design***.* The habituation and test stimuli were computer-animated events created with Macromedia Director 8.0 for PC. Children were habituated to six events (see Figure 1). In one event, children saw two objects motionless on the screen side-by-side. One object was a red square and the other object was a blue circle, and in the center of the object was either a white cross or a yellow heart. A second habituation event was identical to the first one except the left-right position of the two objects was reversed. In the other four habituation events, children saw the blue circle and red square as well as a blue pentagon and a red chair-like shape—all without any surface feature—move along one of two motion paths. The motion paths were the same as those used in previous studies (Rakison, 2004; Rakison & Poulin-Dubois, 2002). One of the motion paths was curvilinear, with the object moving up and down twice in each event, and the other motion path was rectilinear, with the object moving up and down four times in each event. The length of time it took each event to be completed was the same for the two motion paths and the stationary events (8.0 s), and each event could be repeated up to three times per trial. Each presentation of an event was separated by a blue screen that descended and ascended over a period of 1s.

---------------------------------------------

Insert Figure 1 about here

---------------------------------------------

All of the habituation trials began with the two stationary events. The order of these initial stationary habituation events was counterbalanced across participants. After these stationary events, the participants were presented with the appropriate four moving events. Participants who first saw the circle with the cross (or heart) and the square with the heart (or cross) were presented with the square and the chair shape moving curvilinearly (or rectilinearly) and the circle and the hexagon moving rectilinearly (or curvilinearly). The order of these four habituation trials with motion was counterbalanced across participants. After the presentation of the first six events, infants were presented with the same block of events—two static and four dynamic—in a counterbalanced order. The block of habituation events could be repeated up to 3 times in total, for a maximum of 18 habituation trials.

Following habituation, children were presented with two test trials, both of which included a novel pink muffin-shaped object that moved in the same way as one of the habituation stimuli (see Figure 1). Note that infants as young as 4 months of age can discriminate red from pink (Franklin & Davies, 2004). In the *consistent test* event, the muffin-shaped object had the surface feature that was previously paired with one of the static objects, and it moved on the path that was seen for that object during habituation. For example, if children were habituated to the static red square with a white inner cross and the red square and red chair-like shape that moved rectilinearly, in the test phase they would see the muffin-shape with the white cross move rectilinearly. In the *inconsistent test* event, the muffin-shaped object had the surface feature that was previously paired with a static object, and it moved on the path that was not seen for that object during habituation. For instance, if subjects were habituated to the static red square with a white cross and the red square and red chair-like shape that moved rectilinearly, in the test phase they would see the muffin-shape with the white cross move curvilinearly. The pairing of inner shapes and motion paths during the test phase and the order of the consistent and inconsistent test trials was counterbalanced across subjects in each age group.

**Apparatus and procedure.** Each child sat on their caretaker’s lap in front of a computer screen (size: 14 in. x 24 in.; distance: 35 in.) in a small, quiet, softly lit laboratory room. During the habituation and test phase, each event was presented on a computer monitor until the child looked away from the monitor for over 1 s or after 30 s of continuous looking. The habituation phase ended when a child’s looking time for a block of three trials decreased to 50% of that registered during the first three trials or until 18 trials total were presented. A green expanding and contracting circle on a black background with a synchronous bell sound was presented before the first habituation trial and between each habituation and test trial to reorient participants to the computer monitor. The experiment was controlled by HabitX (Cohen, Atkinson, & Chaput, 2000) on an Apple G4 computer.

**Coding and analyses.** The length of the infant’s looking times was coded online by an experimenter’s key press and recorded by the computer. A second judge independently recoded the length of looking on every trial for a random 25% of the children in each age group. The correlation between the second coder’s results and the original coder’s results and the mean difference between the two were calculated. Reliability for children’s visual fixations in the two experiments presented here was *r* > .98, and the mean difference between the two judges on each trial was less than .3 s.

**Results**

An initial set of analyses examined the rate of habituation for the two age groups for both total looking time and number of trials required to reach criterion. The first analysis revealed that 20-month-olds (*M* = 8.79, *SD* = 2.47) and 26-month-olds (*M* = 9.08, *SD* = 3.50) required an equivalent number of trials to reach habituation, *t*(46) = 0.33, *p*>.7. Note that, on average, subjects in both groups were presented with at least six trials and therefore saw both the static and the dynamic events. The second analysis revealed that although the 26-month-olds (*M* = 113.96, *SD* = 60.16) required on average 16 seconds longer to reach criterion than the 20-month-olds (*M* = 97.19, *SD* = 40.84), this difference was not reliably different, *t*(46) = 1.13, *p*>.2.

---------------------------------------------

Insert Figure 2 about here

---------------------------------------------

The primary analysis compared the mean looking times of the two age groups during the two test trials. The looking times are presented in Figure 2. The data were entered into a 2 (test trial: consistent vs. inconsistent) × 2 (age: 20 months vs. 26 months) mixed design analysis of variance (ANOVA). The analysis revealed no significant main effect of test trial, *F*(1,46) = .125, *p>.*7, and no significant effect for age, *F*(1,46) = 0.55, *p>.*4. However, there was a significant interaction between age and test trial, *F*(1,46) = 12.20, *p*<.001, partial *η2* = .21. Planned comparisons revealed that infants at 20 months of age looked longer at the inconsistent test trial (*M* = 10.13, *SD* = 7.13) than the consistent test trial (*M* = 6.04, *SD* = 3.49), *F*(1,23) = 5.73, *p*<0.025, partial *η2* = .20. In contrast, the 26-month-olds looked significantly longer at the consistent test trial (*M* = 11.05, *SD* = 6.88) than the inconsistent test trial, (*M* = 6.54, *SD* = 5.62), *F*(1,23) = 7.45, *p<.*025, partial *η2* = .25.

A final set of analyses compared children’s looking on the last three habituation trials (averaged) to the two test trials. The rationale for this analysis was to determine whether infants in the test trials recovered visual attention relative to the habituation trials, which is particularly relevant for the current design because both test trials included a novel object. The analyses revealed that 20-month-olds looked significantly longer at the inconsistent test trial than the average of the last three habituation trials, *t*(23) = 3.49, *p*<.005, but they looked equally long at the consistent test trial and the last three habituation trials, *t*(23) = .92, *p*>.3. In contrast, the 26-month-olds looked significantly longer at the consistent test trial than the average of the last three habituation trials, *t*(23) = 3.21, *p*<.005, but they looked equally long at the inconsistent test trial and the last three habituation trials, *t*(23) =.12, *p*>9. These analyses suggest that longer looking to either of the test trials did not result from the introduction of the novel stimulus during the test events. In other words, if participants increased looking to the test trials based on the novel shape alone, they would have been expected to demonstrate an increase in looking to both test trials relative to the last habituation events.

**Discussion**

The results of Experiment 1 suggest that 20- and 26-month-olds are able to learn correlations among features that are not presented simultaneously, and that they exhibit the same pattern of learning demonstrated by 7- to 11-month-olds in Yermolayeva and Rakison (2016). Specifically, the 20-month-olds looked longer at the inconsistent test trial relative to the consistent test trial and the 26-month-olds looked longer at the consistent test trial relative to the inconsistent test trial. Longer looking to one of the test trials relative to the other one could only have resulted from second-order correlation learning during the habituation trials because participants were presented with all aspects of the stimuli (i.e., the different shapes, features, and motions) equally during habituation and the novel stimulus—that is, the muffin-shaped body—was involved in each test event. Moreover, subjects must have formed a representation during habituation that captured the relation between the inner shape of the objects and the movement of those objects because these were the only two features of the habituation events that were presented during the test trials.

Although the data suggest that both age groups engaged in second-order correlation learning, an important question is why 20- and 26-month-olds demonstrated a different pattern of preference during the test trials. This pattern of preference is the same as that observed in previous work on second-order correlation learning in 7- to 11-month-olds (Yermolayeva & Rakison, 2016), yet the cause of this switch in infants’ pattern of preference remains unknown. Experiment 2 presents a computational model that was generated to provide insight into this developmental trajectory.

**Experiment 2**

The goals of the following simulation were (a) to model the developmental pattern observed in Experiment 1, and (b) to explore the possible mechanism for this pattern of behavior. In particular, we sought to simulate the findings from Experiment 1 and better to understand the mechanism that underpin these findings. It is worth noting that the findings from Experiment 1 were also obtained by Yermolayeva and Rakison (2016) using a different methodological paradigm with 7- and 11-month-olds, and is therefore unlikely to be an artifact of the design used in Experiment 1 reported here. Thus, the primary aims of the model were to demonstrate that a connectionist associative learning mechanism can form second-order correlations and examine the developmental time course underlying the emergence of a consistent and inconsistent test preference.

**Method**

**Network architecture.** We used a three-layer autoencoder network that was trained using backpropagation as the weight-modification learning rule and momentum (Doug's Momentum: <http://tedlab.mit.edu/~dr/Lens/Commands/dougsMomentum.html>). Connectionist autoencoder networks learn to reproduce the pattern of activation along the input units on the output units. An autoencoder was used here because—similar to infants in a habituation paradigm—these networks learn through error correction iteratively to align their internal representation with the incoming visual input (for a discussion, see Mareschal, French, & Quinn, 2000). Simulations were conducted using the Lens neural-network simulator (Rohde, 1999). Learning rate, momentum, and weight decay were set to 0.08, 0.9, and 0.001, respectively. Note that we ran four additional simulations to determine to what extent the results reported below were robust to adjustments to different learning parameters. In addition to replicating the results reported below with the original parameter settings, we made the following adjustments to the parameters: (1) The number of hidden units was changed from 15 to 13; (2) the number of hidden layers, each with 15 units, was changed from 1 to 2; and (3) weight decay was changed from .001 to .005. Twenty networks—each initialized with a fresh set of small, random weights—were run for each simulation (see Table 1 for results). It can be seen that none of the changes to the parameters altered the qualitative pattern of the below results reported below.

The input to the network consisted of patterns of activity across four groups of input units: a shape, path, color, and feature input group. The shape group consisted of distributed activity across ten units, whereas activity was encoded locally across two units for the path and feature groups and across three units for the color group (see Figure 3). Shape was encoded in a distributed manner to reflect the representational similarity between different shapes in the real world. Path, color, and feature were represented orthogonally (i.e., locally), in contrast, because it is believed that they are the principal perceptual contrasts over which networks—and by extension, infants—learn to form second-order correlations. Each input group projected to one, 15-unit hidden layer, which in turn projected to output groups that corresponded to each input group. Note that we have no reason to believe that our results would have been qualitatively different than if the latter three groups were represented in a different way. That we successfully simulated infants' performance in Experiment 1 and generated predictions that were confirmed in Experiment 3 reinforces this point and suggests that the model presented here is a viable one of infants' second-order correlational learning.

**Materials.** The stimuli used to train and test the network served as rough proxies to those used in Experiment 1. Thus, networks were trained using stimuli that were either red or blue, that contained either an internal cross shape or heart shape, and that moved curvilinearly or rectilinearly (see Figure 3). To equate the network’s training experience with that of infants in Experiment 1, networks were trained on four stimuli that were intended to correspond with the four shape stimuli in Experiment 1 (discussed below in the Training section).

---------------------------------------------

Insert Figure 3 about here

---------------------------------------------

**Training**

A total of 100 networks—each taken to represent a new subject and where 50 of these networks corresponded to “younger infants” who were trained for approximately 200 epochs (*M* =203.78) and the remaining 50, to “older infants” who were trained for approximately 500 epochs (*M* = 564.59)—were trained for this simulation. During each training epoch, the weights—which reflect the network’s knowledge about the correlations between object features—are adjusted to reduce the discrepancy between the desired and actual output. We chose to train the younger networks for approximately 200 epochs and the older networks for approximately 500 epochs for two reasons. First, repeated simulations showed that inconsistent- and consistent-test preferences emerged at reliably different times; that is, repeated simulations showed that an inconsistent preference consistently emerged around 200 epochs, whereas a consistent preference consistently emerged around 500 epochs. Second, we chose to train older networks for more epochs than younger networks given (a) that the information-processing abilities of 26-month-olds are greater than those of 20-month-olds (e.g. Rakison & Poulin-Dubois, 2001) and (b) that older infants presumably expend more cognitive resources—based on such improved information-processing abilities—than younger infants in tasks such as that used here. Each “subject” was simulated by restarting the network and initializing a fresh set of small random weights. Similar to Experiment 1, each network was trained initially to predict an internal cross shape on the output when given as input a red object that did not move and an internal heart shape when given a blue object that did not move. In the second phase —in which internal shape features were absent—to the network had to predict rectilinear (described above in Experiment 1) when given a red square and red chair-like shape and curvilinear motion when given a blue circle and blue pentagon.

**Testing**

At test, the network was presented with a novel muffin-shaped pink object that was presented as a new 10-unit distributed pattern with a familiar internal feature and had to predict on the output the correct motion that corresponded to that feature. The test events—which were identical to those in Experiment 1—included a consistent test event and an inconsistent test event; the same novel pink muffin-shaped object participated in both events. Given that networks were trained to learn that things with internal cross shapes moved rectilinearly and that things with internal heart shapes moved curvilinearly (counterbalanced), the consistent test event corresponded to the event in which the novel pink shape with the internal cross shape moved rectilinearly. In contrast, the inconsistent test event corresponded to the event in which the novel pink shape with an internal cross shape moved curvilinearly.

Networks that correctly predicted rectilinear motion when given the novel pink object with an internal cross shape or curvilinear motion when given the same object with an internal heart shape were said to have a “consistent preference. In contrast, networks that incorrectly predicted rectilinear motion when presented with the novel pink object with a heart shape or curvilinear motion when given the same object with a cross shape were said to have an “inconsistent” preference. Thus, if there was greater activation in the rectilinear output unit than in the curvilinear output unit (and that activation exceeded the commonly accepted threshold of 0.5 [e.g., McClelland & Thompson, 2007]), the network was said to have a "consistent" preference. In contrast, if there was a greater activation that exceeded 0.5 in the curvilinear output unit than in the rectilinear output unit, the network was said to have an "inconsistent" preference. Note that greater activity in one unit compared to another was interpreted as a coarse proxy for longer “looking” to one of the two units.

**Results**

All analyses were conducted in R (R Development Core Team, 2008). In line with the analyses conducted in Experiment 1, mean looking times to the inconsistent and consistent test events for the older and younger networks were examined. We fit separate linear mixed-effects models (LMMs) to the data. This represents a better approach than either univariate ANOVA or ordinary least squares regression because LMMs better address unbalanced and non-independent designs and data (for an extended discussion, see Baayen, Davidson, & Bates, 2008). Thus, we fit an LMM to the mean looking times to the inconsistent and consistent test events for the older and younger networks, where subjects were included as a random-effect factor and where test trial (inconsistent vs. consistent) and age (younger networks vs. older networks) were included as fixed-effects factors. The results revealed no significant main effect of age, *F*(1, 98) = 3.01, *p* = .08. However, there was a significant main effect of condition, *F*(1, 98) = 10.95, *p* < .005, which was qualified by significant interaction between age and test trial, *F*(1, 98) = 2187.25, *p* < .0001.Follow-up planned comparisons revealed that younger networks “looked” significantly longer at the inconsistent test trial (*M* = .733, 95% Bootstrapped CI[.72, .75]) than at the consistent test trial (*M* = .302, 95% CI[.29, .31]), p < .0001 (*n* = 10,000), whereas older networks looked significantly longer at the consistent test trial (*M* = .73, 95% Bootstrapped CI[.72, .74]) than at the inconsistent test trial (*M* = .28, 95% Bootstrapped CI[.32, .35]), *p* < .00001 (*n* = 10,000). As can be seen in Figure 4 below, the data from the simulation presented here qualitatively match those presented in Figure 2 from infants in Experiment 1.

To examine whether a preference for the inconsistent and consistent test events emerged at reliably different times for the younger and older infants, we fit a separate LMM to the data, where consistency (inconsistent vs. consistent) was treated as the sole fixed-effects factor, whereas subjects were treated as the random-effect factor. This analysis revealed a significant main effect of consistency effect, *F*(1, 98) = 457.65, *p* < .0001, which indicated that longer looking to the inconsistent test event (*M* = 203.78, 95% Bootstrapped CI[192.77, 213.75]) emerged reliably earlier than to the consistent test event (*M* = 564.59, 95% Bootstrapped CI[544.16, 586.54]). It is important to mention that it is an open question whether and in what ways these differences map onto different developmental ages.

---------------------------------------------

Insert Figure 4 about here

---------------------------------------------

**Discussion**

The results of the simulation demonstrate that—like the 20- and 26-month-olds in Experiment 1—younger networks “looked” longer at the inconsistent test than the consistent test event, whereas older networks “looked” longer at the consistent test event than the inconsistent test event. These findings notwithstanding, it is unclear why an inconsistency preference emerged prior to a consistency preference. One potential answer is suggested by the weights in the younger networks.

As can be seen in the top left panel of Figure 5, the largest weights are those in the color, feature, and motion clusters but not in the shape cluster. Generally, weights in connectionist models are taken to represent the strength of the associative relation between features (e.g., McClelland & Thompson, 2007), where larger weights correspond to stronger relations and darker and lighter weights correspond to negative and positive weights, respectively. Thus, the fact that the weights in the color, shape, and path units are larger than the weights in the shape cluster suggests that the younger networks had learned the associative relation between these features but not between shape and these features. This is because shape varied to a greater extent during training, which required greater information-processing abilities to process, which the younger networks did not possess after approximately 200 epochs of training. Thus, the reason 20-month-olds preferred the inconsistent test event to the consistent test event presumably was because the second-order relation between feature and motion path was violated in this event but not in the consistent test event.

---------------------------------------------

Insert Figure 5 about here

---------------------------------------------

To understand why a preference for the consistent test event emerges later than one for the inconsistent test event, it is necessary to compare the relative magnitudes of the weights in the color, feature, and path groups in the younger networks to those in the older networks. As can be seen in the left and middle panels in Figure 5, the weights in the color, path, and feature clusters were smaller in the older networks than in the younger networks, whereas the weights in the shape cluster were noticeably larger in the older networks than in the younger ones. This implies that the reason the older networks “looked” longer at the consistent test event than at the inconsistent test event was because the weights in the color, feature, and path groups were insufficiently strong—compared to those in the younger network—for older networks (and, by loose extension, 26-month-olds) to detect a violation in the relation between path and feature but not so weak that a representation of these features was not present at all (as would be the case if there was no activity in the color, feature, and path groups). This means that the presentation of the familiar features at test, coupled with an incomplete representation of the first- and second-order relations, caused the network preferentially to attend to the consistent test event but not to the inconsistent test event. This explanation is consistent with that of Hunter and Ames (1988) who noted that infants will show a familiarity preference when their representation of the input is present—as is the case in this simulation by the small, but not negligible, weights—but requires more processing to be fully encoded.

The distribution of weights in the older networks predicts that in the absence of shape variability, 26-month-olds would fail to encode any of the dimensions and thus should look equally long at both test events because the color, feature, and path clusters would not have to compete with the shape cluster for representational resources. In terms of the 20-month-olds, the model predicts that in the absence of shape variability they will attend solely to and encode the shape dimension—because this is the least variant cue compared to the color, path, and feature dimensions—and thus should look longer at both test events relative to the last few habituation trials because both test events introduce a novel shape. We tested these predictions in Experiment 3.

**Experiment 3**

In this experiment, 20- and 26-month-olds’ ability to learn second-order correlations was tested in a non-category context; that is, participants were habituated to only three of the events used in Experiment 1. Thus, subjects were shown two stationary objects (one red and one blue) that possessed surface features and then shown each of those objects as they moved along one of two motion paths. The question of interest was whether participants could use second-order correlation learning to determine which surface feature should move along which motion path when presented with information about a single object that embodied that correlation and to test the predictions of the PDP model presented in Experiment 2.

---------------------------------------------

Insert Figure 6 about here

---------------------------------------------

**Method**

**Participants***.* The participants were twenty healthy full-term 20-month-olds (mean age 20 months 2 days; range = 19;15 to 20;14) and twenty healthy full-term 26-month-olds (mean age 26 months 0 day; range = 25;05 to 26;16). There were 11 females and 9 males in the 20-month-old group and 9 females and 11 males in the 26-month-old group. The majority of participants were White and of middle socioeconomic status. Data from an additional eighteen 20-month-olds were excluded from the final sample, 13 because of fussing, 1 because of parent interference, and 4 because of technical problems and experimenter error. Data from an additional twenty-three 26-month-olds were excluded from the final sample, 9 because of fussing, 4 because of parent interference, 2 for failure to habituate, 4 for displaying test trial looking times that were more than 2 standard deviations away from the mean, and 4 because of technical problems and experimenter error. Although this number of dropouts is high, it is not unusual for experiments that test young children with the habituation paradigm (e.g., Cohen & Oakes, 1993; Rakison & Poulin-Dubois, 2002). Subjects were recruited in the same way as Experiment 1.

**Stimuli, Design, Apparatus, and Procedure.** The stimuli were the same as those in Experiment 1, but there were only three habituation events (the two static objects, the circle moving, and the square moving; see Figure 6). All other aspects of the experiment were identical to Experiment 1.

**Results**

As in Experiment 1, the initial analyses compared the rate of habituation for the two age groups for both total looking time and number of trials required to reach criterion. The analyses indicated that 20-month-olds (*M* = 9.56, *SD* = 4.14) and 26-month-olds (*M* = 7.90, *SD* = 2.86) required the same number of test trials to reach habituation, *t*(38) = 1.47, *p*>.1. They also showed that although the 20-month-olds (*M* = 115.27, *SD* = 59.57) required on average over 20 seconds longer to reach criterion than the 26-month-olds (*M* = 93.21, *SD* = 53.32), this difference was not significantly different, *t*(38) = 1.23, *p*>.2.

---------------------------------------------

Insert Figure 7 about here

---------------------------------------------

The mean looking times of the two age groups during the two test trials are presented in Figure 7. Children’s looking times to the two test events were analyzed with a 2 (test trial: consistent vs. inconsistent) × 2 (age: 20 months vs. 26 months) mixed design ANOVA. The analysis revealed no significant main effect of test trial, *F*(1,38) = 0.75, *p>*.3, η2p = .02, and no significant interaction between test trial and age, *F*(1,38) = 0.47, *p>*.4, η2p = .01. The analysis also indicated that 20-month-olds looked significantly longer at the test trials overall (*M* = 14.31, *SD* = 8.12) than the 26-month-olds (*M* = 7.57, *SD* = 5.79), *F*(1,38) = 17.10, *p*<.001, η2p = .31. This main analysis confirmed the first prediction made by the model for 20- and 26-month-olds in a non-category context; that is, equal looking to the two test trials.

A final analysis compared children’s looking on the last three habituation trials (averaged) to the two test trials to examine the model’s prediction that 20-month-olds should look longer at both test trials relative to the last few habituation trials. The analyses revealed that 20-month-olds looked significantly longer at the inconsistent test trial (*M* = 15.47, *SD* = 8.41) than at the last three habituation trials (*M* = 6.12, *SD* = 2.99), *t*(19) = 5.08, *p*<.0001, and they looked significantly longer at the consistent test trial (*M* = 13.15, *SD* = 7.84) than at the last three habituation trials, *t*(19) = 5.02, *p*<.0001. In contrast, the 26-month-olds looked equally long at the consistent test trial (*M* = 7.33, *SD* = 5.55), *t*(19) = 0.51, *p>*.6, the inconsistent test trial (*M* = 7.60, *SD* = 6.05), *t*(19) = 0.73, *p>*.4, and the last three habituation trials (*M* = 6.54, *SD* = 3.19). These results confirm the predictions of the model and suggest that 20-month-olds may have looked relatively long to both test trials because of the introduction of the novel body shape.

**Discussion**

The results of the current experiment indicated that 20- and 26-month-olds looked equally long at the consistent and inconsistent events following habituation to stimuli that embodied an unobservable correlation between features. This implies that children at this age are unable to infer an unseen correlation between a static and dynamic feature of an object on the basis of two other observed correlations when those correlations are not embedded in a category context; that is, when multiple exemplars exhibit the same motion path. This pattern of data supported the predictions of the computational model presented in Experiment 2; that is, it was predicted that both 20- and 26-month-olds would look equally long at both the consistent and inconsistent test events if the variability of shape in the category context (as in Experiment 1) served as a facilitating cue to direct infants to attend to the more systematic information available in the events (e.g., feature, color, and path). The model also predicted that because of the lack of a facilitating cue, 20-month-olds may have not learned any of the feature correlations in the events. This was borne out by the fact that infants at this age increased looking to both of the test trials relative to the last habituation events, which suggests that the pattern of looking during the test trials was affected primarily by the introduction of a new body shape.

In conjunction with Experiment 1, the present data suggest that 20- and 26-month-olds are able to learn second-order correlations that involve static and dynamic features, but they need to be exposed to multiple exemplars that exhibit that correlation to do so presumably because the variability they encode in a category context better supports generalization to a novel stimulus (Colombo et al., 1990).

**General Discussion**

The main goals of the three experiments presented here were to determine if, when, and how 20- and 26-month-olds can associate two visual features of objects—one static and one dynamic—that were not presented together in a category and non-category context. These experiments are only the second series to investigate whether infants and young children can engage in second-order correlation learning whereby they extend two learned associative links to a third, unseen correlation, and they are the first to do so with dynamic, moving stimuli. Experiments 1 and 3 showed that 20- and 26-month-olds inferred an unseen correlation between a dynamic and a static feature from two observed ones in a category context but not in a non-category context. Experiment 2 presented a connectionist model that provided a mechanistic explanation of Experiment 1—that centered on differing information-processing abilities between younger and older infants—and generated predictions about second-order correlation learning in a non-category context that were tested and confirmed in Experiment 3. In particular, the model predicted—and Experiment 3 confirmed—that both age groups engaged in second-order correlation learning only when presented with multiple instances of one of the correlations (e.g., P is correlated with R1, Q is correlated with R1 and R2; therefore, P is correlated with Q) but not when presented with a single instance of one of the correlations.

The finding that young children are capable of second-order correlation learning—especially in a category context—suggests that it may represent a potentially powerful form of generalization in the first years of life. It could, for instance, help to explain how infants and young children extend their knowledge to novel features, novel objects, and novel category members. There is considerable evidence that infants’ earliest representations are grounded in the perceptual, surface features of objects and correlations among those features. For example, infants in the second year of life associate agency, path of motion, and self-propulsion with the external moving parts of objects (Rakison, 2005a, 2005b, 2006). The results from infants in the current experiments and the computational model reported here extend these findings and show that infants and young children can generalize these associations to other features that are only indirectly related to one of those features through second-order correlation learning; that is, once infants learn that two features are correlated they will start to associate other features—even those that are not directly observed as correlated with one of those features—with them. In this way, over developmental time—and arguably across many domains with structured input—infants and young children construct increasingly rich associatively derived representations for the features and properties of objects and entities in the world.

It is worth noting that subjects in the current experiments and infants in the studies by Yermolayeva and Rakison (2016) behaved somewhat differently in a non-category context. In experiments reported here, subjects showed no preference for either test event in a non-category context whereas in the work by Yermolayeva and Rakison (2016) infants at 7 and 11 months of age demonstrated a preference for one or other test event depending on their age. In our view, learning in both cases relied on the same general associative learning mechanism and therefore this difference in behavior likely resulted from two factors. First, in the current experiments children were presented with dynamic events whereas in Yermolayeva and Rakison (2016) infants were presented with static objects. The former are inherently more complex than the latter and this may have made it more difficult for infants to process the feature correlations among the stimuli. Second, infants in the experiments by Yermolayeva and Rakison (2016) were presented with 3D objects that they could interact with and rotate to inspect—whereas the events in the current experiments were presented on a computer screen—and this may have made it easier for infants to process the feature correlations among the stimuli.

Before closing, it is worth mentioning two potential criticisms of the model presented in Experiment 2. The first limitation was that motion was represented locally—as one of two binary features—rather than dynamically. This is potentially problematic because there is evidence that suggests that motion is difficult to encode because it is often only intermittently available in the input (e.g., Rakison, 2006; Rakison & Poulin-Dubois, 2002). Our decision to represent motion locally rather than dynamically was nonetheless warranted because we sought to simulate the transition from an inconsistency preference to a consistency preference in 20- and 26-month-olds rather than to focus on the effects of motion of second-order correlation learning, per se. Second, we did not equate exactly the number of training epochs and infants' real-world age. Despite the fact that this may be problematic, this is a general criticism of developmental connectionist models (e.g., Cohen, Chaput, & Cashon, 2002; Rakison & Lupyan, 2008). These criticisms notwithstanding, the simulation presented here is valuable because it demonstrated that second-order correlation learning can be an emergent ability and it generated predictions that were tested and confirmed in Experiment 3.

In sum, the current experiments and simulation demonstrate that 20- and 26-month-olds are capable of second-order correlation learning whereby associations between two pairs of features (e.g., P and Q, P and R) are generalized to the features that were not presented together (e.g., Q and R). We propose that second-order correlation learning is grounded in associative processes and may represent the origins of the kinds of deductive reasoning that is observed in older children (Markovits, 1993; see also Sloman, 1996), and as such it represents a relatively unexplored process that may support infants’ and young children’s ability to generalize their existing knowledge to novel features, objects, and events across a wide range of domains.

**References**

Barr, R., Marrott, H., & Rovee-Collier, C. (2003). The role of sensory preconditioning in memory retrieval by preverbal infants. *Learning and Behavior, 31*, 111–123.

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*(2), 105-141.

Cohen, L.B., Atkinson, D.J., & Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition. (Version 2.2) [Computer Software]. Austin, TX. The University of Texas at Austin.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development*, *17*(3), 1323-1343.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology*, *29*(3), 421.

Colombo, J., McCollam, K., Coldren, J. T., Mitchell, D. W., and Rash, S. J. (1990). Form

categorization in 10-month-old infants. Journal of Experimental Child Psychology, 49, 173-188.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological review*, *112*(2), 347-382.

Cuevas, K., Rovee-Collier, C., & Learmonth, A. E. (2006). Infants form associations between memory representations of stimuli that are absent. *Psychological Science, 17,* 543-549. Dias, M. G., & Harris, P. L. (1988). The effect of make-believe play on deductive reasoning. *British Journal of Developmental Psychology, 6*, 207-221.

Dewar, K. & Xu, F. (2010) Induction, overhypothesis, and the origin of abstract knowledge: evidence from 9-month-old infants.  *Psychological Science, 21*, 1871-1877.

Dwyer, D.M., Mackintosh, N.J., & Boakes, R.A. (1998). Simultaneous activation of the representations of absent cues results in the formation of an excitatory association between them. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 163–171.

[Fiser, J., & Aslin, R.N. (2001)](http://www.bcs.rochester.edu/people/aslin/pdfs/Fiser_Aslin2001.pdf). Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychological Science, 12, 499-504.

Franklin, A., & Davies, I. R. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, *22*(3), 349-377.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3-to 4-month-old infants: simulations and data. *Journal of Experimental Psychology: General*, *133*(3), 382-397.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology, 69,* 133-149.

Harris, P.L., Kavanaugh, R.D., & Dowson, L. (1997). The depiction of imaginary transformation: Early comprehension of a symbolic function. *Cognitive Development, 12,* 1–19.

Holyoak, K. J., & Morrison, R. G. (2005). Thinking and reasoning: A reader’s guide. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning (pp.1-9)*. Cambridge, UK: Cambridge University Press.

Hunter, M. A. & Ames, E. W. (1988). A multifactor model of infant preferences for novel and

familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research,*

Vol. 5. (pp. 69-95). Norwood, NJ: Ablex.

Hull, C.L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

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*(5), e36399.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35-B42.

Madole, K. L., & Cohen, L. B. (1995). The role of object parts in infants' attention to form-function correlations. *Developmental Psychology, 31*, 637-648.

McClelland, J. L., & Thompson, R. M. (2007). Using domain‐general principles to explain children's causal reasoning abilities. *Developmental Science*, *10*(3), 333-356.

Marcus, G. (2001). The algebraic mind: Integrating connectionism and cognitive science. Cambridge, MA: Bradford Books.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental psychology*, *36*(5), 635.

Mandler, J. M., & McDonough, L. (1996). Drinking and driving don't mix: Inductive generalization in infancy. *Cognition*, *59*(3), 307-335.

Markovits, H., Schleifei; M., & Fortier, L. (1989). Development of elementary deductive reasoning in young children. *Developmental Psychology, 25*, 787-793.

McClelland, J. L., & Thompson, R. M. (2007). Using domain‐general principles to explain children's causal reasoning abilities. *Developmental Science*, *10*(3), 333-356.

Mou, Y., Province, J. M., & Luo, Y. (2014). Can infants make transitive inferences? *Cognitive Psychology 68,* 98–112.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: toward an adaptive process account of successes and failures in object permanence tasks. *Psychological review*, *104*(4), 686.

Pavlov, I.P. (1927). *Conditioned reflexes*. London: Oxford University Press.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463-475.

R Development Core Team (2008). R: a language and environment for statistical computing. R Foundation for Statistical Computing,Vienna,Austria.URLhttp://www.R-project.org.

Rakison, D. H. (2004). Infants’ sensitivity to correlations among static and dynamic features in a category context. *Journal of Experimental Child Psychology, 89*, 1-30.

Rakison, D. H. (2005a). A secret agent? How infants learn about the identity of objects in a causal scene. *Journal of Experimental Child Psychology, 91*, 271-296.

Rakison, D.H. (2005b). Developing knowledge of motion properties in infancy. *Cognition, 96,* 183-214.

Rakison, D. H. (2006). Make the first move: How infants learn about the identity of self-propelled objects. *Developmental Psychology, 42*, 900-912.

Rakison, D. H., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology*, *34*(1), 49-62.

Rakison, D. H., & Cohen, L. B. (1999). Infants’ Use of Functional Parts in Basic‐like Categorization. *Developmental Science*, *2*(4), 423-431.

Rakison, D.H., & Poulin-Dubois, D. (2002). You go this way and I’ll go that way: Developmental changes in infants’ attention to correlations among dynamic parts in motion events. *Child Development, 73,* 682-699.

Rakison, D.H. & Lupyan, G. (2008). [Developing object concepts in infancy: An associative learning perspective.](http://www.sas.upenn.edu/%7Elupyan/rakison_lupyan_inPress.pdf) [*Monographs of the Society for Research in Child Development,*](http://www.blackwell-synergy.com.proxy.library.cornell.edu/toc/mono/73/1) *73*, 1-110.

Rohde, D. L. (1999). LENS: The light, efficient network simulator. School of Computer Science, Carnegie Mellon University, 1-8.

Sandoval, M., & Gomez, R. (2013). The development of nonadjacent dependency learning in natural and artificial languages. *Wiley Interdisciplinary Reviews: Cognitive Science 4 (5)*, 511-522.

Slater, A., Mattock, A., Brown, E., & Bremner, J. G. (1991). Form perception at birth: Revisited. *Journal of experimental child psychology*, *51*(3), 395-406.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926-1928.

Tomasello, M., Striano, T., & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology, 17*, 563–584.

Walker, C. M., & Gopnik, A. (2014). [Toddlers infer higher-order relational principles in causal learning](http://journals.sagepub.com/doi/abs/10.1177/0956797613502983). *Psychological Science, 25*, 161-169.

Yermolayeva, Y., & Rakison, D. H. (2014). Connectionist modeling of developmental changes in infancy: Approaches, challenges, and contributions. *Psychological bulletin*, *140*(1), 224.

Yermolayeva, Y. & Rakison, D. H (2016). Seeing the unseen: Second-order correlation learning in 7- to 11-month-olds. *Cognition, 153*, 87-100.

Younger, B.A., & Cohen, L.B. (1986). Developmental change in infants’ perception of correlations among attributes. *Child Development, 57*, 803-815.

Younger, B. A., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology, 24*, 611-619.

Table 1

EFFECTS OF CHANGING PARAMETERS

|  |  |  |  |
| --- | --- | --- | --- |
|  | Replication Simulations | Bayes Factors[[1]](#footnote-1) | |
| Replication with the original parameters | *F*(1, 18) = 352.06, *p* < .0001 | | BF10 > 5,000 |
| 13 hidden units | *F*(1, 18) = 289.34, *p* < .0001 | | BF10 > 5,000 |
| 2 hidden layers (15 hidden units each) | *F*(1, 18) = 85.25, *p* < .0001 | | BF10 > 5,000 |
| .0005 weight decay | *F*(1, 18) = 179.02, *p* < .0001 | | BF10 > 5,000 |

Figure Captions

Figure 1. Examples of stimuli used in Experiment 1. Subjects saw the two static events first, followed by the four dynamic events. These six events could be presented up to three times each during habituation. There were two test events, both with a novel shape: the inconsistent test event paired the surface feature from one of the static objects with an inappropriate motion path; the consistent test event paired the surface feature from one of the static objects with an appropriate motion path.

Figure 2. Mean looking time and standard errors to the two test trials in Experiment 1.

Figure 3. A simple schematic of the autoencoder network used to simulate the results of Experiment 1. The four input groups are connected to the two hidden layers which in turn are connected to the four output groups. The network’s job was to recreate the pattern of activation in each input group on each of the corresponding output groups.

Figure 4. Comparison the mean looking time to the two test trials in Experiment 1 and the mean error produced by the network to the two test trials.

Figure 5. A diagram of the weights in the two hidden layers and four output groups (i.e., shape, color, feature, and path output groups) in the younger network. The left and middle panels depict the weights that were produced once the younger and older network showed a preference for the inconsistent and consistent test events, respectively. The right panel depicts the weights that were produced once the preference for the consistent test event emerged and the preference for the inconsistent test event reemerged in even older networks. The relative size and color of weights corresponds to whether the weight is positive (white) or negative (black). Output groups that are highlighted by a red circle represent the features to which the networks attended most following approximately 200, 500, and 600 epochs of training.

Figure 6. Example stimuli from Experiment 3. The design was identical to Experiment 2 except that the features were on the outside of the objects instead of the inside.

Figure 7. Mean looking time and standard errors to the two test trials in Experiment 3.

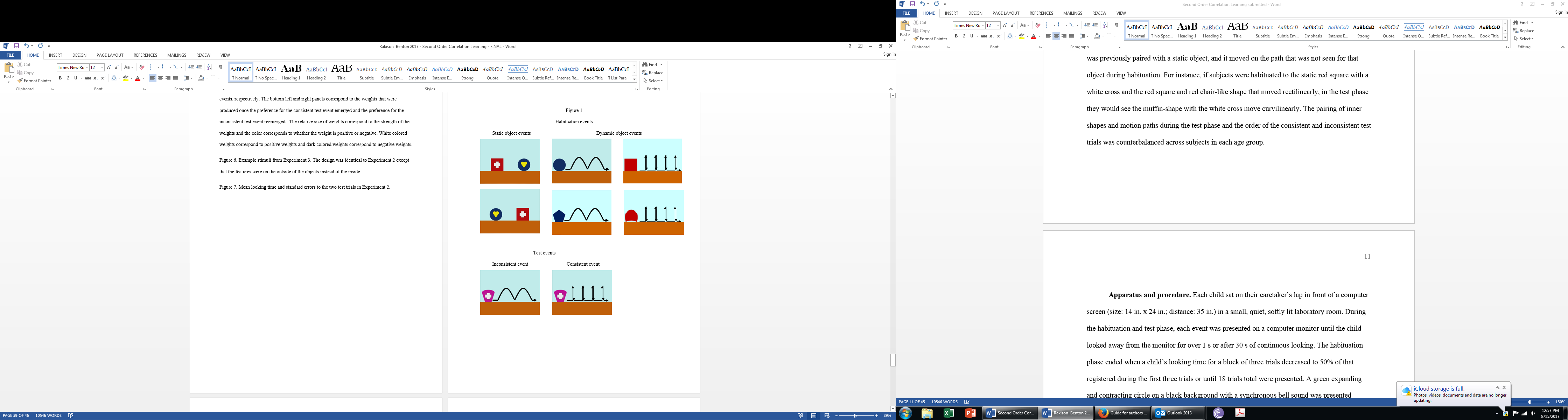
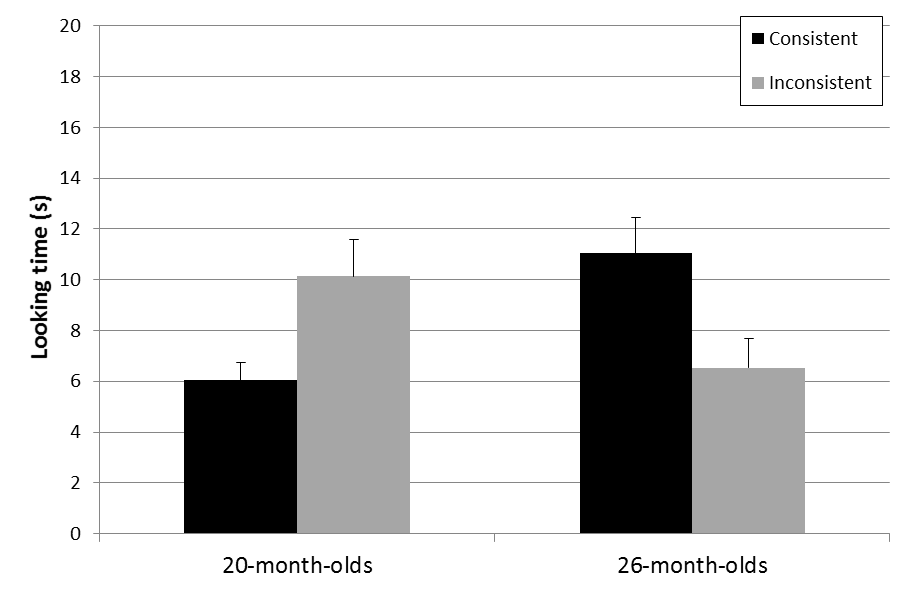
Figure 1

Figure 2



*p*<.025

*p*<.025

Figure 3

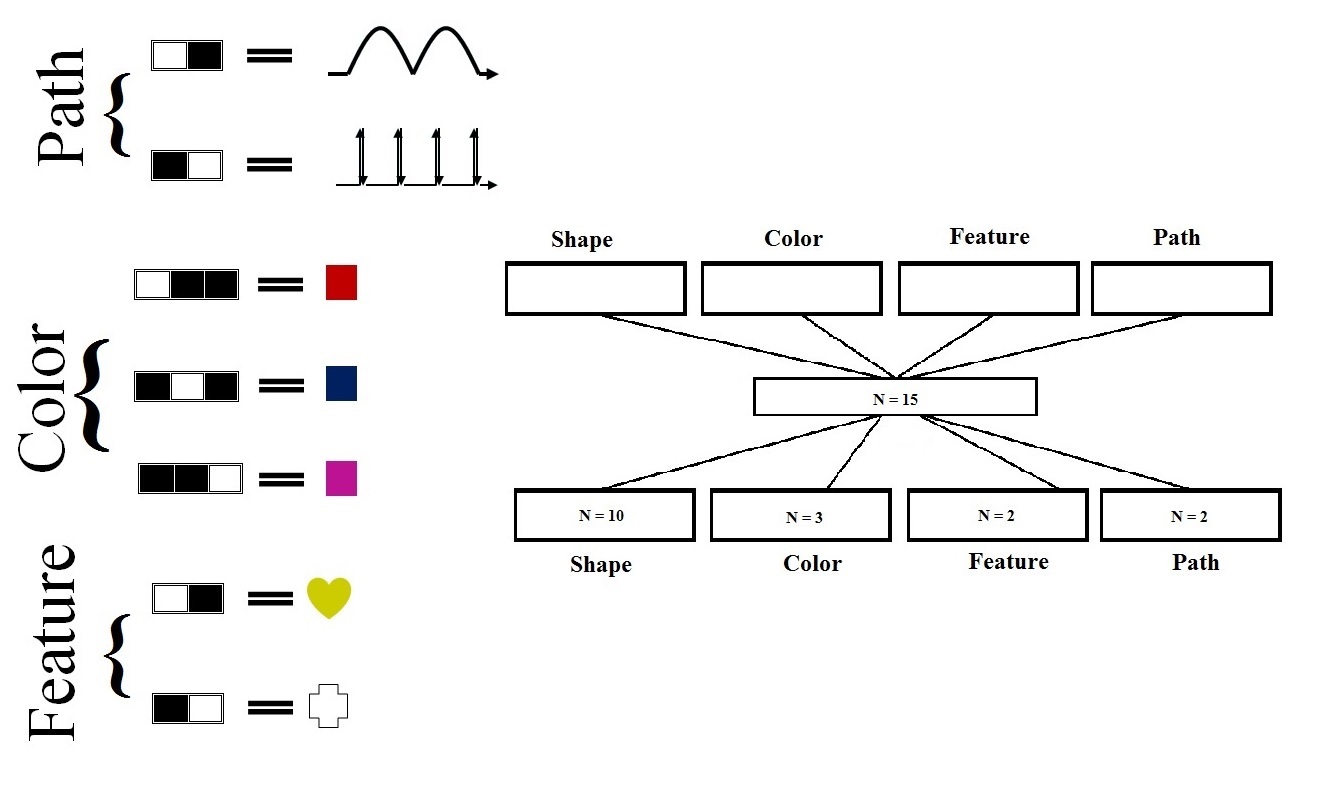
****

Figure 4

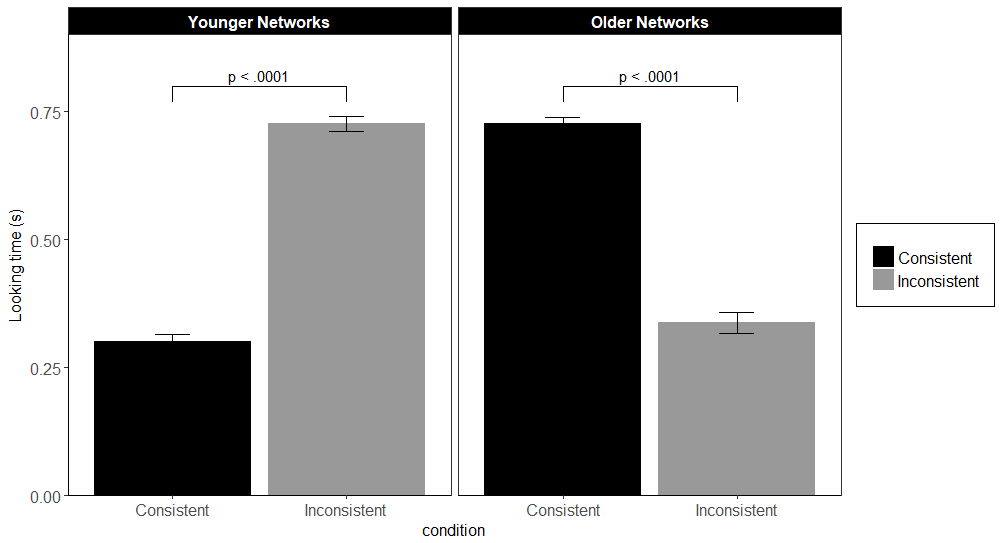


Figure 5

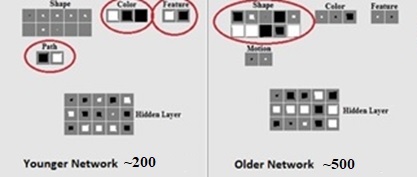
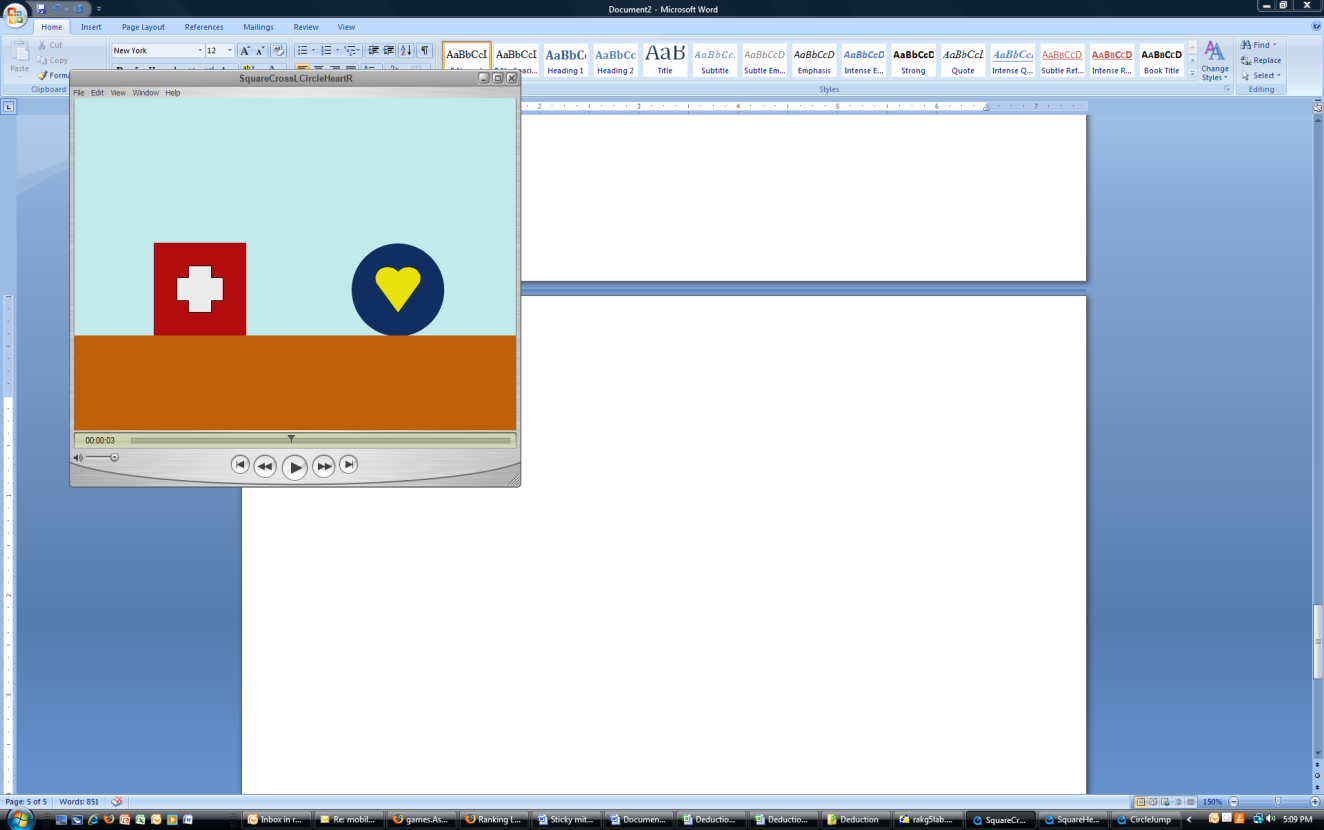
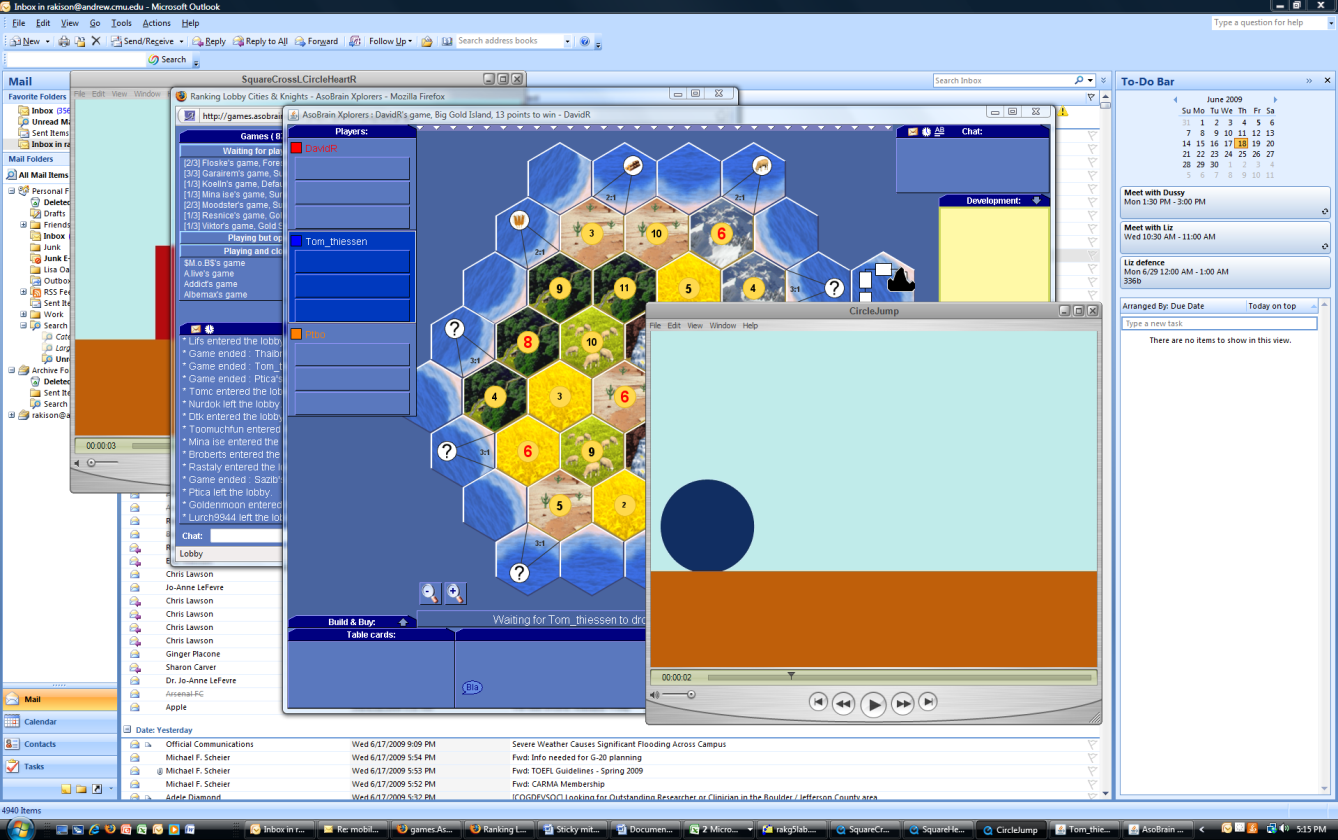
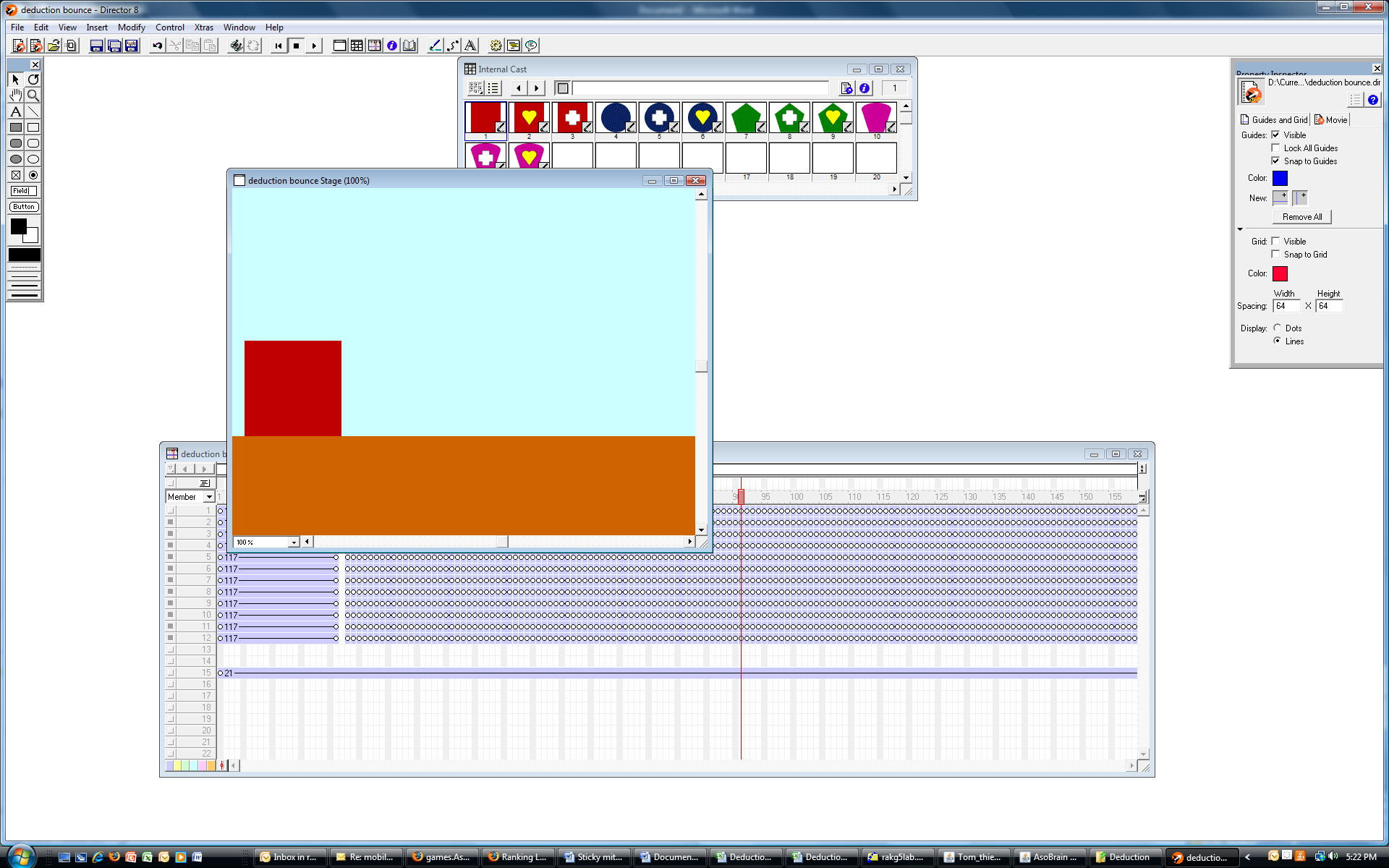


Figure 6

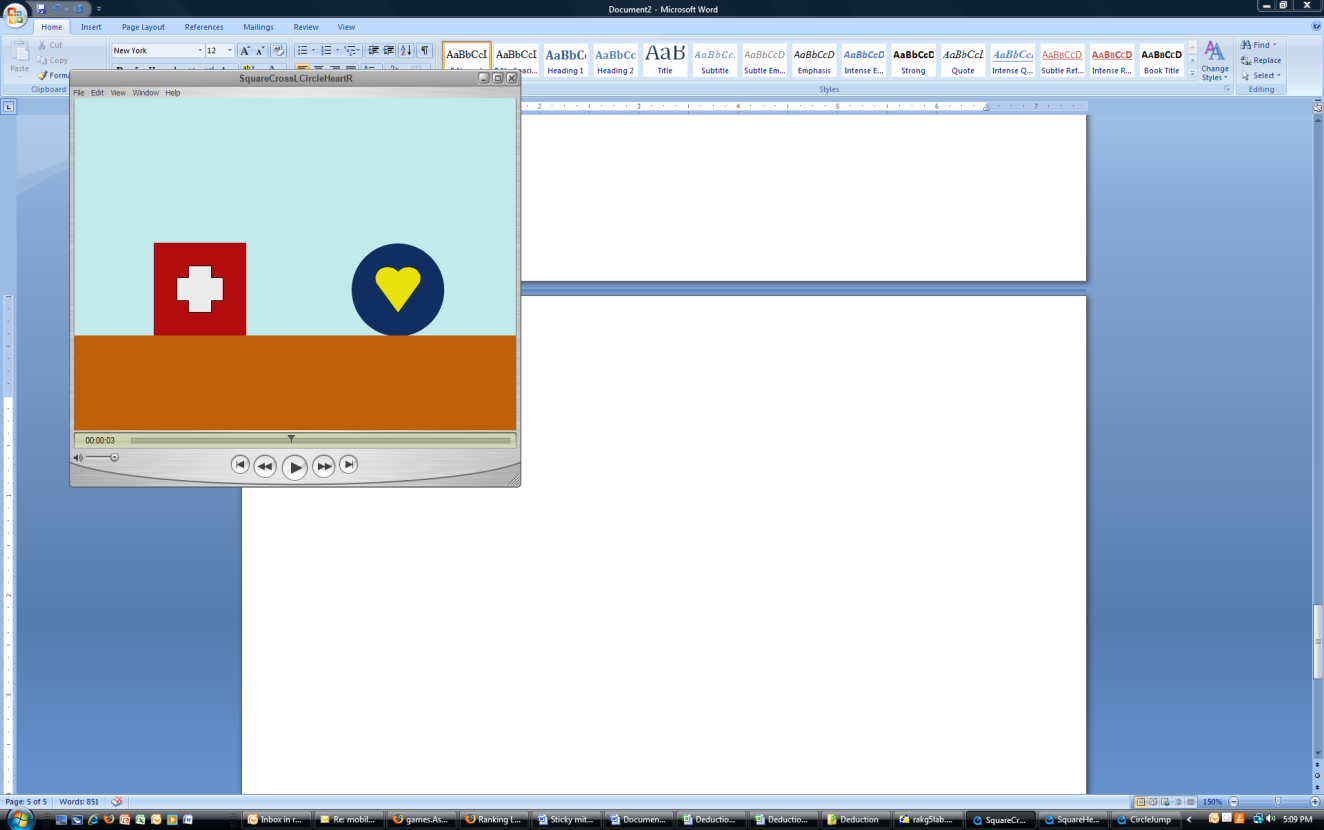
Test events

Inconsistent event

Habituation events



Consistent event



Static object events

Dynamic object events

Figure 7

1. Note that the data were also estimated using a Bayes Factor using Bayesian Information Criteria (Wagenmakers, 2007) in which the fit of the data under the alternative hypothesis is compared to that under the null hypothesis; that is, we estimated BF10 for each simulation under the different parameter settings. This analysis indicated that the data were 5,000 times more likely under the alternative hypothesis than under the null hypothesis for each of the four simulations. Using Raferty’s (1995) and Jeffrey’s (1961) guidelines for interpreting Bayes Factors, this indicates very strong and decisive evidence in support of the alternative hypothesis for each of the simulation runs. [↑](#footnote-ref-1)