INFANCY
THE OFFICIAL JOURNAL OF THE
INTERNATIONAL CONGRESS
OF INFANT STUDIES

Infancy, 1–15, 2018

ISSN: 1525-0008 print / 1532-7078 online
DOI: 10.1111/infa.12225

WILEY Blackwell

Limits on Composition of Conceptual Operations in 9-Month-Olds

Steven T. Piantadosi, Holly Palmeri, and Richard Aslin
Department of Brain and Cognitive Sciences
University of Rochester

Complex systems are often built from a relatively small set of basic features or operations that can be combined in myriad ways. We investigated the developmental origins of this compositional architecture in 9-month-old infants, extending recent work that demonstrated rudimentary compositional abilities in preschoolers. Infants viewed two separate object-occlusion events that depicted a single-feature-change operation. They were then tested with a combined operation to determine whether they expected the outcome of the two feature changes, even though this combination was unfamiliar. In contrast to preschoolers, infants did *not* appear to predictively compose these simple feature-change operations. A second experiment demonstrated the ability of infants to track two operations when not combined. The failure to compose basic operations is consistent with limitations on object tracking and early numerical cognition (Feigenson & Yamaguchi, *Infancy*, 2009, *14*, 244). We suggest that these results can be unified via a general principle: Infants have difficulty with multiple updates to a representation of an unobservable.

Readers of Lewis Carroll's *Through the Looking Glass* have no trouble understanding Humpty Dumpty's idea of an "*un*birthday present," a gift given to a friend because today is *not* their birthday. This concept can easily be grasped because it is the composition of two already existing concepts: *un-* and *birthday*, which can be straightforwardly combined to create a representation denoting the days that are not one's birthday. This type of conceptual combination reflects one of the most powerful aspects of human cognition, allowing adults to fluidly create novel concepts throughout many domains including language, music, mathematics, and complex motor actions. As a result, such compositionality has been at the heart of some of the earliest cognitive theories (Boole, 1854; Frege, 1892), and foundational debates about the core properties of thought (Chalmers, 1993; Fodor, 1975; Fodor & McLaughlin, 1990; Fodor & Pylyshyn, 1988; Smolensky, 1988, 1998; Van Gelder, 1990).

Correspondence should be sent to Steven T. Piantadosi, Department of Brain and Cognitive Sciences, 358 Meliora Hall, P.O. Box 270268, University of Rochester, Rochester, NY 14627-0268. E-mail: spiantado@gmail.com

Despite the importance of compositionality, little work has empirically investigated its origins in human development. In particular, it is not known how early learners have access to compositionality as a tool for building new representations—at what age may concepts be productively combined? At the same time, compositionality has often been assumed in computational models as the core generative mechanism for sophisticated conceptual structures, such as knowledge of number, magnetism, or word meanings (Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Goodman, Ullman, & Tenenbaum, 2009; Katz, Goodman, Kersting, Kemp, & Tenenbaum, 2008; Kemp, Goodman, & Tenenbaum, 2008; Piantadosi, 2011; Piantadosi & Jacobs, 2016; Piantadosi, Tenenbaum, & Goodman, 2012; Siskind, 1996; Ullman, Goodman, & Tenenbaum, 2010). Such models allow learners to operate over a computationally rich hypothesis space by positing that complex computations are built by composing simpler operations.

In this compositional framework, concepts have their origin in a set of "built-in" primitives that are common to all human learners and from which all other concepts are ultimately derived. Of course, there is substantial debate about what these primitives are, ranging from rather rich representations, to a handful of specialized domains (Spelke & Kinzler, 2007) or operations (e.g., Chomsky, 1995), to quite rudimentary biases for attention and learning (Aslin & Newport, 2012; Endress & Bonatti, 2007). What is undeniable is that even the most radical empiricist must posit a set of primitives, such as reward, comparison, or memory decay, to account for the acquisition of concepts.

Recent work with preschoolers (Piantadosi & Aslin, 2016) showed that by age 3.5–4.5 years, children *are* able to predictively compose two novel operations. Children (mean age 50.9 months, range 42.9–53.9) were trained on two individual operations, represented as object occluders that changed the features of objects passing behind them. After reaching a training criterion for each single-feature operation, they were tested on their predictions about the outcome of objects that passed behind both occluders—undergoing both operations—without being visible between occluders. This task used a touch screen interface where children had to select what the object would look like when it came out from two occluders, choosing out of four possible options. We will use the mathematical notation of *functions* to describe these operations and their composition. Thus, in the training phase, preschoolers saw two separate functions, $f(x)$ and $g(x)$, and in the testing phase, they saw a novel composition of functions, $f(g(x))$.

Preschoolers' performance in this task was well-above chance but also far from ceiling: They gave the correct answer to two binary feature changes about 50% of the time, with a chance rate of 25%. However, this accuracy on predicting $f(g(x))$ was about the same as their accuracy on learning just $f(x)$ and $g(x)$ alone. Performance was influenced by training time and accuracy, but not by demographic predictors such as age and sex. These results indicate that by approximate 4-years-old children are able to accurately predict the outcome of compositions of functions after having been trained only on the pieces.

In the present series of studies, we extend these investigations of preschoolers' knowledge of conceptual composition to infants to investigate the developmental origins of this ability. The sense of *compositionality* that we consider is analogous to the case of preschoolers, where infants are shown two operations separately and their looking patterns are used to index what they expect when those two operations are combined, with the outcome of the first not visible until the second has applied. The experiments therefore ask whether infants' representation of a latent outcome (e.g., $g(x)$) can *then* be used as input to a second function, f, forming $f(g(x))$. Note that in general, this ability will not be independent of their representational and memory capacities for these functions. We show that at age 9 months of age, infants do not show behavior consistent with correct predictions of function compositions. Experiment 1 demonstrates that infants show looking time patterns consistent with implicit belief that only the second of two functions applied to the object. In mathematical notation, they act as although the composition $f(g(x)) = f(x)$. In Experiment 2, we show that this failure is likely *not* due to an inability to track two operations ($f(x)$ and $g(x)$), simultaneously. Infants exhibit looking patterns consistent with learning two separate object transformations, suggesting that their difficulty lies in the process of combination itself.

EXPERIMENT 1: INFANTS FAIL TO CORRECTLY COMPOSE FUNCTIONS

Infants were presented with two boxes that each depicted an operation on a moving object. Figure 1a shows the setup of the experiment. This display contained four elements: a novel object for each trial (here, a knobby circle, on the left), a primary occluder (long gray rectangle) and two "operators" (boxes with red and dotted icons) which "paint" their features onto any object that passes behind the occluder while the operator icon was present. The leftmost (red) operator is considered the "first" because it is the first that the novel object passes behind when it moves to the right.

In a single trial, infants observed the first (here red) operator slide onto the long occluder. The object then travelled (here from the left) behind the occluder and

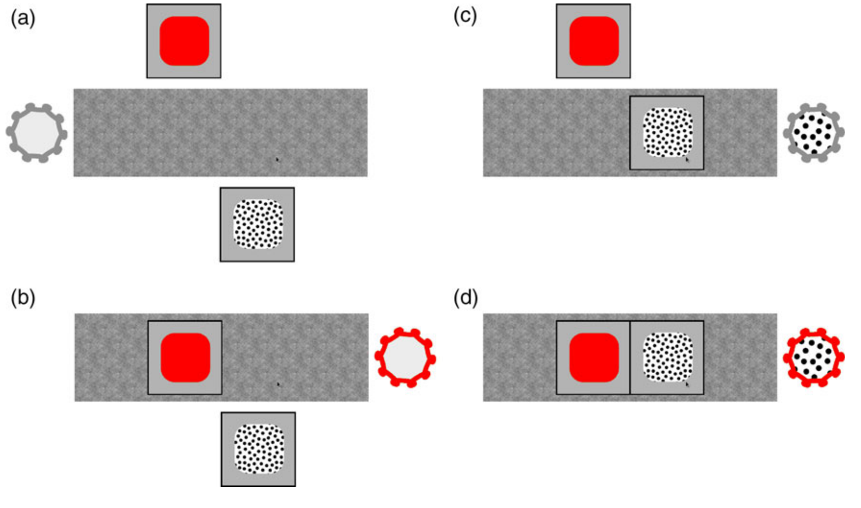

Figure 1 Stimuli for Experiment 1. An object appears to the left of a long occluder (a). Infants are shown that when the "red" box is in place, it changes the object's border color (b). When the "dotted" box is in place, it adds dots to the center (c). After familiarization with these two operations, the test of compositionality is what outcome infants expect when both boxes apply (d).

emerged on the right with the operation applied (Figure 1b), where it wiggled briefly. Next, they observed the red operator slide off the occluder into its starting position, and the dotted operator slide into place on the occluder. A new object appeared on the left, the dotted operator moved into place on the occluder, and the object went behind the occluder, emerging on the right with dots (Figure 1c). The large gray occluder (Figure 1a) ensures that the moving object is hidden for the same amount of time regardless of how many operators are "active."

These displays were used to familiarize infants to each of the two operators immediately before each critical test trial. The key display consisted of both operators sliding into position on the occluder (Figure 1d), suggesting that *both* operators should apply. Infants observed the object pass behind both operators and appear with one of three possible outcomes: Only the first operator has applied ("First," same outcome as in 1 (b)), only the second operator has applied ("Second," same outcome as in 1(c)), or both operators have correctly applied ("Correct," shown in 1(d)). Each of these displays continued to show the object moving back and forth with the same outcome until infants terminated attention by looking away from the display for more than 2 sec. The amount of time spent looking at these three critical test trials was the dependent measure used to infer implicit knowledge of compositionality.

Figure 2 shows hypothetical outcomes for each of the three test conditions if infants based their expectations about the object's transformation behind the occluder on one of three implicit hypotheses: (a) that only the first operator applied, (b) that only the second operator applied, or (c) that both operators applied (i.e., the correct composition). Under the assumption that infants will look the longest to unexpected events and the least to expected events, Figure 2 illustrates that the most expected outcome is the "First" operator in (a), the "Second" operator in (b), and the "Correct" combined operators in (c). In addition, because the combined operators include a feature change consistent with either the "First" or the "Second" operator, infants who only applied a single operator should find the "Correct" combined operators as less unexpected than the wrong single operator. For example, if the two operators were red and dots and you saw red+dots as the outcome, but you expected only red as the outcome because it was the first operator, then you would be less surprised by red+dots than by dots alone (without red). By determining which possible pattern of outcomes (a–c) in Figure 2 most closely captures infant behavior on the critical test trials, we can infer

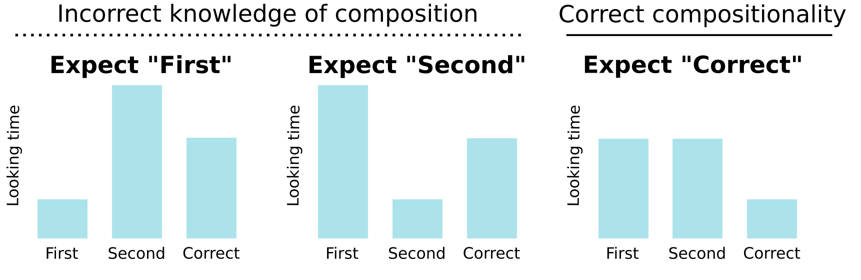

Figure 2 Predicted looking time outcomes under the assumption that looking times are shortest to expected outcomes. Each subplot corresponds to a logically possible expectation infants might have about observed compositional trials: (a) expecting only the first function to apply in compositional displays, (b) expecting only the second, or (c) expecting the correct composition.

which types of outcomes they expect from compositional displays. In this way, the experiment tests compositionality against strong alternative expectations (i.e., first or second operators), and failures of compositionality result in statistically nonnull patterns of looking times.

Participants

A total of 33 infants were run in Experiment 1. Trials that terminated before the critical object was revealed were removed as the participants did not observe the critical outcome. This resulted in effective removal of two subjects due to zero remaining data. Beyond these cases, no infants were removed due to fussiness or failure to attend to the stimuli. Overall, this resulted in an average of 5.35 critical trials for each infant.

This study was conducted according to Declaration of Helsinki guidelines, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this study were approved by the Research Subjects Review Board at the University of Rochester.

Methods

After each infant arrived in the lab, they were seated on their parent's lap approximately 48 inches in front of a 42-inch wide-screen plasma monitor. Displays consisted of simple animations created and displayed in Kelpy, a free and open source kid experimental library in Python (Piantadosi, 2012)[1]. For each infant, the particular color, pattern, and order of these operations were fixed throughout the experiment. Soft instrumental accordion music[2] was played during the experiment to help sustain attention and interest.

Individual infants were shown only two functions (e.g., "red" and "dots," as in Figure 1a). To keep the number of conditions low, we manipulated the outcome (Correct/Function 1/Function 2) partially across subjects, such that each infant either saw Correct and Function 1 as the possible outcomes, or Correct and Function 2. This kept the number of conditions to two for each infant. Trials were presented in blocks of the two conditions. The specific operations and their order of application were consistent across trials for each infant, and randomized across infants. Each trial used a different randomly chosen object from a set of four simple geometric shapes.

Infant look-aways from the display were monitored by video and coded in real time using Kelpy by lab staff blind to the experimental condition and current display. Lookaways more than 2-sec terminated the trial, at which point an attention-getter was redisplayed before a new trial began. Sessions were terminated when infants demonstrated a loss of interest in the experiment.

Results

Looking times in the experiment exhibited significant outliers, ranging between 0.08 and 61.94 sec (mean = 11.71*s*). To handle this, outliers were removed using a repeated

[1]Experimental code is available from the first author.

[2]Composed by Yann Tierson for the movie *Am'elie*.

Grubbs test (Grubbs, 1950) in each condition, removing 9% of the total trials collected.

Figure 3 shows infants' mean (bars) and median (triangle) looking times to each critical test condition, aggregating by subject. Individual subject means are shown as black dots. This figure shows that 9-month-old infants exhibit their lowest looking time when the outcome is the second function. Under a simple linking function where increased looking times correspond to greater "surprise," the results indicate that infants expect that the combination of two operations will yield an outcome where only the second one has applied.

We analyzed the data using a mixed-effect linear regression (Gelman & Hill, 2007) with main effects of trial and pattern-versus-color, as well as random intercepts by subject, the maximal random effects structure justified by the data. This analysis first reveals an overall (omnibus) main effect of condition, indicating that looking times vary significantly between the three critical test conditions ($\chi^2_2 = 18.2$, $p = .0001$). The looking time to the second function outcome is significantly less than to both the first function outcome ($\beta = 10.7$, $SE = 2.4$, $t = 4.5$, $p < .001$) and the correct outcome ($\beta = 5.7$, $SE = 2.0$, $t = 2.87$, $p = .005$). Additionally, looking times to the correct

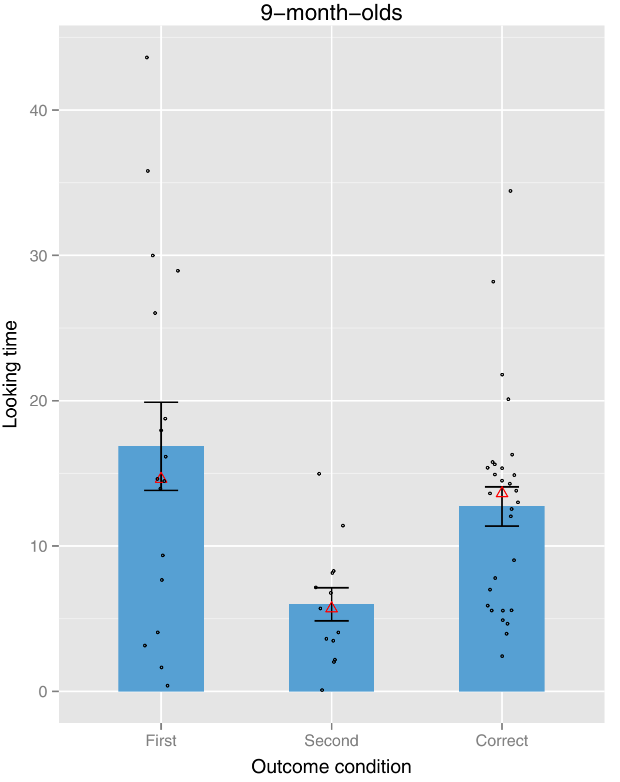

Figure 3 Infant looking times *after* the critical object reveal when both operators are applied (i.e., Figure 1d). Bars show condition means (mean of subject means) and standard errors on the raw data, points show individual infants, and triangles show the median looking times. Note that the error bars shown here are conservative as they do not reflect within-infant aspects of the design.

outcome are significantly less than those to Function 1 ($\beta = 5.0$, $SE = 2.1$, $t = 2.4$, $p = .02$).

This regression also included an effect of (standardized) trial number, which was statistically significant ($\beta = -2.6$, $SE = 0.8$, $t = 3.1$, $p = .002$) with a trend such that infants look less for later trials in the experiment, likely indicating fatigue. Additional regression analyses revealed no effects of whether the first operator manipulated color or pattern ($\beta = 1.8$, $SE = 1.7$, $t = 1.0$, $p = .32$).

Discussion

These results strongly suggest that infants do not automatically infer the correct composition of color and pattern feature-transformations. Instead, their looking time patterns are consistent with expecting the *second* function only to have applied to an object. Under the assumption that looking times relate simply to degree of nonexpectation, the rankorder of all bars in Figure 3 correspond to the "Expect Second" prediction shown in Figure 2. The "correct" outcome shares some features with the expectation of Function 2, as it has the expected dots in the middle (it also has an unexpected red border). The Function 1 outcome is least expected as it has none of the predicted features: a red border and *no* dots in the middle. Thus, while our findings represent a negative result with respect to the presence of compositionality, they are not statistically null and they show interpretable patterns according to the three possible predictions.

It is important to point out; however, that the relationship between looking times and expectation may *not* be monotonic. Kidd, Piantadosi, and Aslin (2012) present quantitative evidence for a U-shaped relationship between the stimulus "surprisal" and looking time. Thus, it is possible that any of the outcomes depicted in Figure 2 might be the "most expected" depending on where these conditions lie on the U-shaped curve. We note this as a theoretically interesting caveat to our preferred interpretation; namely that infants are biased to utilize only a single, final operator when an object undergoes a transformation. The U-shaped model is difficult or impossible to evaluate in the current situation because it requires a statistical model of expectation,[3] which would be hard to construct for these displays without first knowing infants' compositional abilities. Additionally, it is possible—as in nearly all infant experiments—that low-level factors such as the perceptual complexity or difference from the starting object could drive increased looking times, particularly to the "correct" outcome. Our results are therefore suggestive of failure of compositionality, contingent only on our assumptions about how looking behavior relates to perception and prediction.

One potential concern with interpreting infants' failure to compose operations is that this may not be reflective of a compositional limitation *per se*, but could simply reflect an inability to manipulate and represent multiple functions. For instance, infants may have difficulty representing multiple object transformations regardless of whether or not they are composed. Perhaps infants can only track one transformation of objects, and their behavior on compositional displays simply reflects that limitation. In general, we believe this interpretation is unlikely because infants do not show nonsensical or null looking patterns as might be expected with catastrophic failures of representation (see, e.g., Feigenson & Carey, 2005). Instead, they exhibit looking patterns

[3]Kidd et al. (2012) used an ideal statistical learner of sequences.

matching a consistent expectation about the displays. However, to more strongly rule out an overall inability to represent two functions, Experiment 2 examines infants' ability to learn and represent two functions simultaneously, without composition.

EXPERIMENT 2: INFANTS CORRECTLY REPRESENT TWO SEPARATE FUNCTIONS

Experiment 2 trained infants on two functions corresponding to the simple object feature changes used in Experiment 1 (color/pattern changes). Instead of testing on the composition of operations, infants were tested only on the individual functions (without combination) after the single-operation familiarization phase (see Figure 4). Success on this task would indicate that infants' failure in Experiment 1 is not due to limitations of tracking two object transformations, but rather to the process of combination itself.

Participants

Thirty-two 9-month-old infants were run in Experiment 2 (mean age: 9.5 months, range 9.0–10.0 months; 17 females). Trials were removed where infants did not see the critical outcome (19% of trials), and outliers were again trimmed using a repeated Grubbs' test (5% of remaining trials), resulting in the effective removal of two infants.

This study was conducted according to Declaration of Helsinki guidelines, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects in this

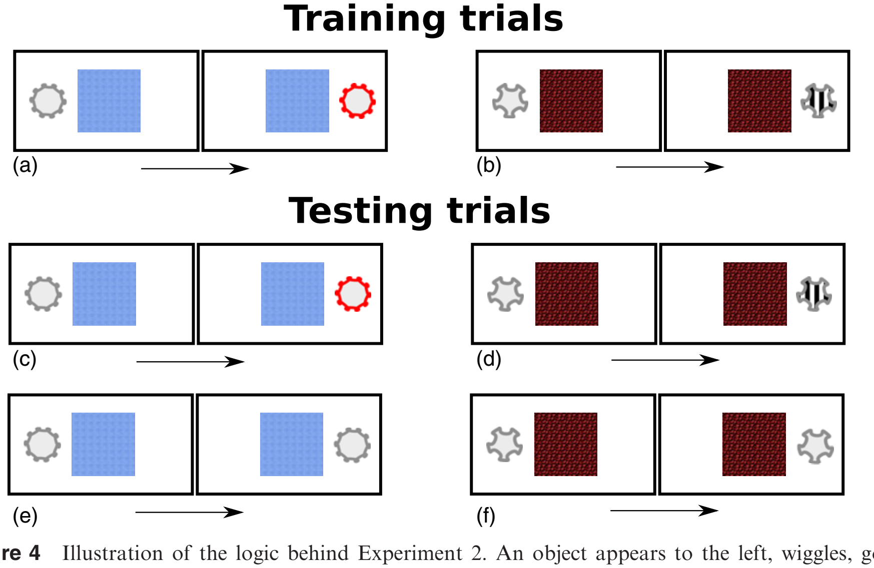

Figure 4 Illustration of the logic behind Experiment 2. An object appears to the left, wiggles, goes behind an occluder, and emerges with a featural change (a, b), with training blocks randomly intermixed with each transformation type. The occluder cued whether the change affected the object's pattern or color. Testing trials in (c–f) evaluated expectations that some change would occur when objects passed behind each occluder.

study were approved by the Research Subjects Review Board at the University of Rochester.

Methods

Like Experiment 1, Experiment 2 showed infants two transformations: One object went behind the occluder and changed color, the other object went behind the occluder and changed pattern (Figure 1a). Unlike Experiment 1, the occluders used in this experiment did not have iconic patterns cuing the transformation. Instead, infants learned the operation performed by each occluder only by observing it act on a specific object. This provides a stronger test of infants' ability to represent arbitrary pairs of transformations during occlusion. Infants were first familiarized to the transformation performed on each object. To do this, they were shown two familiarization blocks in which each object was altered in appearance (color or pattern) 1-, 2-, and 3-times (presenting a total of 2 blocks $\times$ 3 trials $\times$ 2 objects = 12 total observations).

After this familiarization phase, infants observed up to five test blocks. In each test block, they were shown one correct trial for each transformation (cued by the occluder) and one incorrect trial for each transformation consisting of no transformation for each of the two transformations. This gave a total of four trials per block, balanced between those observed in familiarization and no change. These test blocks measured the expectations infants formed during familiarization: Would they distinguish "Correct" (expected) and "Incorrect" (unexpected) outcomes for both operators, after the brief familiarization phase? In both the training and test blocks, infants only saw objects pass behind a single occluder and they were never tested on compositions of functions.

The key dependent measure was infants' relative looking to the correct and incorrect outcomes in the switch (i.e., no change) blocks, after observing the familiarization trials. If infants are able to track both operators (e.g., object A results in the addition of red, object B results in the addition of pattern), they should distinguish these expected results from the unexpected result of no change in the object as it passed behind the occluder.

Results

The key comparison in Experiment 2 is between outcomes that are expected according to familiarization with those that are unexpected. Figure 5 shows these looking times, demonstrating a difference between expected and unexpected outcomes across the two operators (color and pattern). This difference is statistically significant under a mixed-effect linear regression using by-subject intercepts, the maximum random effects structure justified by the data ($\beta = -1.96$, $SE = 0.8$, $t = -2.47$, $p = .01$). The regression found a significant effect of trial ($\beta = -2.56$, $SE = 0.42$, $t = -6.1$, $p < .001$), as well as a significant effect of whether or not the occluder manipulated a color or pattern ($\beta = 2.08$, $SE = 0.93$, $t = 2.25$, $p = .02$), with longer overall looking times to pattern change outcomes.

For power, this primary analysis analyzed both operations (Color-versus-Pattern). However, it is also important to note that statistically significant results could in this analysis be due to averaging individual infants, each of whom only tracks a single operator. It is difficult to estimate each infant's behavior on both operators, as this

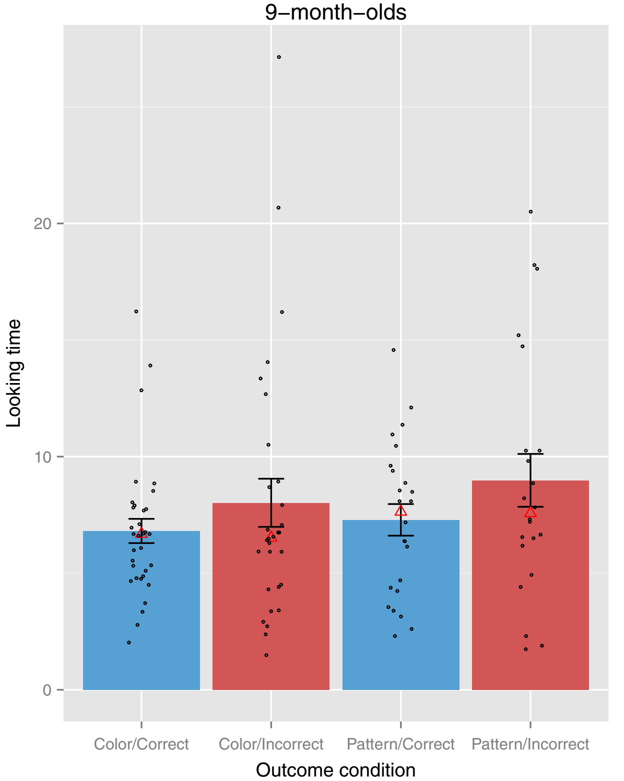

Figure 5 Nine month old looking times to expected and unexpected outcomes in Experiment 2. Bars show condition means (mean of subject means) and standard errors on the raw data, points show individual infants, and triangles show the median looking time. Note that the error bars shown here are conservative as they do not reflect within-infant aspects of the design.

would require comparison of conditions within each individual. However, we can test if there is statistical evidence for variability within individuals between conditions: Does each individual infant treat conditions differently? Our results show no hint of such a pattern, as tested by adding a factor for the operation being observed as an effect for each subject ($\chi^2_2 = 0.18$, $p = .91$).

Similarly, there is no interaction between condition and occluder ($\beta = 2.6$, $SE = 1.7$, $t = 1.5$, $p = .13$), indicating that the effect of expected/unexpected outcomes does not depend on which operator is observed. These null results must be interpreted with care, but they do fail to provide evidence that infants differ between the two operators.

Discussion

These results suggest that infants' failure to compose in Experiment 1 is not due to an overall inability to track two functional operations. Experiment 2 provides evidence that they can track two operations individually, as long as they are not asked to form expectations about their combination. However, it is important to note that the quality of their representation is not pinpointed by Experiment 2. For instance, infants could have been only representing the fact that there was a change caused by each box; determining exactly what they represent about such featural changes will likely require a large series of experimental studies and controls. Nonetheless, representation of only the fact that there was a change predicts a different pattern of looking on Experiment 1: If so, we would have expected uniform looking to all of the outcomes as each includes a change. Thus, when the results of Experiments 1 and 2 are interpreted jointly, the findings are most consistent with the hypothesis that infants fail to compose, yet can form some kind of expectations about what single boxes do to objects.

These results also extend prior results on infants' object and feature tracking (Leslie, Xu, Tremoulet, & Scholl, 1998). Kaldy, Blaser, and Leslie (2006) found that 6-month-olds notice a color but not luminance change to an occluded object, perhaps sometimes remembering an object without some of its features (Kibbe & Leslie, 2011). Our results here indicate that 9-month-olds can track that two changes occur, as well as operators that make those changes. In itself, this result is interesting in that it suggests that such operations can be acquired through observation throughout the course of a short experiment. This point is reminiscent of a distinction in computer science between languages that can manipulate functions like ordinary ("first class") objects (e.g., Scheme), and those that cannot (e.g., C++). Infants appear to be more like the former, able to in principle represent and manipulate not just objects, but operations themselves that transform other objects.

GENERAL DISCUSSION

The results of the two experiments reported here are the first to explore the origins and limitations on conceptual compositionality in the first year of life. Although we found evidence that 9-month-olds can keep track of two independent operations during a series of occlusion events, they are apparently unable to combine these two operations in novel ways. However, we must note several caveats with generally interpreting these negative results. First, our experiments probed one specific type of compositionality: The behavior of operators that change visible features of an object. Many of the learning models that motivated this research use compositionality as a mechanism to generate hypotheses for a statistical learner. It is possible that learners are able to implicitly combine concepts to explain data, while still failing the more explicit, predictive type of compositionality observed here. However, these results are not straightforwardly consistent with the strongest flavors of compositional cognitive theories, in which function combination is active from a very young age, providing a powerful and productive system for creating new representations.

It is interesting to compare these results to Piantadosi and Aslin (2016)'s study with similar stimuli on 3–4-year-olds. There, children's incorrect responses were focused on trials where the same dimension changed twice—that is, where one occluder changed the pattern feature and the second one changed it back. Moreover, when children made errors, they were more likely to give incorrect responses that matched the *second* occluder, meaning that they often acted as although the first occluder did not apply ($f(g(x)) = f(x)$). This pattern is the same as the one we propose to explain the present infant looking patterns and thus or and might reflect a general tendency or limitation that interferes with compositional reasoning in these tasks, persisting in some through the preschool years.

It remains to be seen what change after 9 months of age allows preschoolers to generally succeed on similar tasks. If compositional reasoning is roughly in place by 3–4 years, it is plausible that it depends on other developing abilities—for instance, attention, compositional language, or working memory. There is evidence that 6-month-old infants face a short-term memory limitation of perhaps only a single object (Káldy & Leslie, 2005; Pelphrey et al., 2004; Ross-sheehy, Oakes, & Luck, 2003), with capacity increasing substantially over infants' first year (Ross-sheehy et al., 2003). It may be that the observed compositional limitations could reflect a rational strategy in the face of the substantial cognitive limitations in early infancy.

More broadly, these results extend previous findings on infants' limited ability to update representations. Infants' failures are somewhat reminiscent of Piaget's classic Anot-B error (Piaget, 1954) in which infants fail to search in one location for an object after it has been hidden several times in a different one (see Marcovitch & Zelazo, 1999; Wellman, Cross, Bartsch, & Harris, 1986). More specifically to our paradigm, several prior studies demonstrating similar limitations can be framed as failures of function composition. Wynn (1992) originally demonstrated infants' arithmetic abilities with very small sets (see also Simon, Hespos, & Rochat, 1995): Infants who saw an object added or removed from a small, occluded array of objects correctly expected the correct number of objects when the occluder was removed. If infants form a representation x of the occluded array, this result suggests they are able to update x to $f(x)$ where f is, for instance, a function that adds a single element to x, assuming that the cardinalities involved are sufficiently small (Feigenson & Carey, 2005). However, this updating process appears to be additionally limited by the number of operations performed. Despite the ability of even 5-month-olds to compute $1 + 1$ and $2 - 1$, other results have shown that 7-month-olds (Moher, Tuerk, & Feigenson, 2012) and 10-month-olds (Baillargeon, Miller, & Constantino, 1994) fail on $1 + 1 + 1$, which requires updating a representation two times ($f(f(x))$) after the initial object has been occluded.[4] Analogous to our Experiment 2, infants' failure in these numerical tasks is not about simply performing two operations, but rather performing two operations on the *same* representation: Uller, Carey, Huntley-Fenner, and Klatt (1999) showed that infants succeed if the two object additions are to different occluders in separate locations. Note that in these studies, although, the operations are a memory update of a number of objects, which may be simpler than our transformations of object properties.

Our findings are closely related to Feigenson and Yamaguchi (2009), who showed that 11-month-olds correctly represented quantities of crackers hidden in buckets when the buckets (A and B) were baited sequentially (e.g., AAB), but not in alternation (ABA) (see also Moher & Feigenson, 2013). Just as in our studies, Feigenson and Yamaguchi (2009) shows that the initial object representation is "overwritten" when accessed the second time. This result is closely related to Káldy and Leslie (2005)'s "memory span of one" in which 6.5-month-old infants could only identify the last of two objects hidden behind an occluder.

Our results fit in the general framework of these findings, except that the critical representational component for our results is not the number of objects—as the

[4]Interestingly, 10- and 12-month-olds do succeed when choosing a bucket with three sequentially baited crackers over one with two (Feigenson, Carey, & Hauser, 2002), suggesting a developmental trend and/or a difference across experimental paradigms.

experiments only involve one object—but the number of *changes* that take place to a single object's features. This symmetry with other work suggests a common source for infant's limitations across these domains: Infants may have trouble updating a representation of an unobservable (e.g., unseen feature, object, set) more than one time. Doing so requires a level of abstraction beyond what is possible under the memory and conceptual limitations of infancy. If this explanation in terms of memory limitations is correct, it means that infants may succeed on versions of compositional tasks that place lighter demands on memory, perhaps by giving explicit or continued cues to an object's place, motion, or features. Exploring these directions should be an important goal for extensions of this work.

CONCLUSION

These results have provided a first exploration of infants' early compositional capacities. In contrast to prior results with preschoolers (Piantadosi & Aslin, 2016), infants appear unable to compose simple functions that alter an objects' features. Our results in Experiment 2 and the specific pattern of looking times in Experiment 1 suggest that this is *not* a failure of representing functions themselves. Rather, infants exhibit a consistent way of combining operations that does not accord with our commonsense notion of composition: they expect only the second function to apply ($f(g(x)) = f(x)$). We hypothesize that these patterns of results stem from fundamental limitations on updating a representation of a single unseen property, set, or object more than one time, a pattern consistent with prior results on infant object tracking and failures in simple numerical tasks.

ACKNOWLEDGMENTS

We'd like to thank Roman Feiman, Celeste Kidd, Noah Goodman, and members of AKlab for helpful discussions about this work. We are grateful to babylab RAs for help scheduling and running subjects in this study. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health under Award Number F32HD070544 and an NIH research grant to R. Aslin and E. Newport (HD-037082). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

REFERENCES

Aslin, R. N., & Newport, E. L. (2012). Statistical learning from acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*, 170–176.

Baillargeon, R., Miller, K., & Constantino, J. (1994). *Ten-month-old infants' intuitions about addition*. Unpublished manuscript, University of Illinois at Urbana, Champaign, IL.

Boole, G. (1854). *An investigation of the laws of thought: On which are founded the mathematical theories of logic and probabilities*. London, UK: Walton and Maberly.

Chalmers, D. J. (1993). *Connectionism and compositionality: Why fodor and pylyshyn were wrong*. Didcot, UK: Taylor & Francis.

Chomsky, N. (1995). *The minimalist program* (Vol. 28). Cambridge, UK: Cambridge University Press.

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*, 247–299.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*, 295–313.

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*, 150–156.

Feigenson, L., & Yamaguchi, M. (2009). Limits on infants' ability to dynamically update object representations. *Infancy*, *14*, 244–262.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work. *Cognition*, *35*, 183–204.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis, Connections and symbols. *Cognition*, *28*, 3–71.

Frege, G. (1892). "On Sense and Reference" ["Über Sinn und Bedeutung"], Zeitschrift für Philosophie und philosophische Kritik, *100*(1892), 25–50.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Goodman, N., Tenenbaum, J., Feldman, J., & Griffiths, T. (2008). A rational analysis of rulebased concept learning. *Cognitive Science*, *32*(1), 108–154.

Goodman, N., Ullman, T., & Tenenbaum, J. (2009). Learning a theory of causality. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2188–2193).

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics*, *21*(1), 27–58.

Kaldy, Z., Blaser, E. A., & Leslie, A. M. (2006). A new method for calibrating perceptual salience across dimensions in infants: The case of color vs. luminance. *Developmental Science*, *9*, 482–489.

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*, 153–177.

Katz, Y., Goodman, N., Kersting, K., Kemp, C., & Tenenbaum, J. (2008). Modeling semantic cognition as logical dimensionality reduction. In *Proceedings of thirtieth annual meeting of the cognitive science society*.

Kemp, C., Goodman, N., & Tenenbaum, J. (2008). Learning and using relational theories. *Advances in Neural Information Processing Systems*, *20*, 753–760.

Kibbe, M. M., & Leslie, A. M. (2011). What do infants remember when they forget? Location and identity in 6-month-olds' memory for objects. *Psychological Science*, *22*(12), 1500–1505.

Kidd, C., Piantadosi, S. T., & Aslin, R. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*, e36399.

Leslie, A. M., Xu, F., Tremoulet, P. D., & Scholl, B. J. (1998). Indexing and the object concept: Developing 'what' and 'where' systems. *Trends in Cognitive Sciences*, *2*(1), 10–18.

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, *70*(6), 1297–1313.

Moher, M., & Feigenson, L. (2013). Factors influencing infants' ability to update object representations in memory. *Cognitive Development*, *28*, 272–289.

Moher, M., Tuerk, A. S., & Feigenson, L. (2012). Seven-month old infants chunk items in working memory. *Journal of Experimental Child Psychology*, *112*, 361–377.

Pelphrey, K. A., Reznick, J. S., Goldman, B. D., Sasson, N., Morrow, J., Donahoe, A., & Hodgson, K. (2004). Development of visuospatial short-term memory in the second half of the 1st year. *Developmental Psychology*, *40*, 836–850.

Piaget, J. (1954). *Construction of reality in the child* (Vol. 20). Abingdon, UK: Routledge.

Piantadosi, S. T. (2011). *Learning and the language of thought*. Unpublished doctoral dissertation, MIT. Available from http://colala.bcs.rochester.edu/papers/piantadosi thesis.pdf

Piantadosi, S. T. (2012). *Kelpy: A free library for child experimentation in python*. Available from https://github.com/piantado/kelpy/
Piantadosi, S. T., & Aslin, R. (2016). Compositional reasoning in early childhood. *PLoS One*, *11*, e0147734.
Piantadosi, S. T., & Jacobs, R. (2016). Four problems solved by the probabilistic Language of Thought. *Current Directions in Psychological Science*, *25*, 54–59.
Piantadosi, S. T., Tenenbaum, J., & Goodman, N. (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition*, *123*, 199–217.
Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development*, *74*(6), 1807–1822.
Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of wynn (1992). *Cognitive Development*, *10*, 253–269.
Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to meaning mappings. *Cognition*, *61*, 31–91.
Smolensky, P. (1988). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, *26*(S1), 137–161.
Smolensky, P. (1998). Connectionism, constituency, and the language of thought. *Cognitive Architectures in Artificial Intelligence: The Evolution of Research Programs*, *2*, 183.
Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.
Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, *14*(1), 1–36.
Ullman, T., Goodman, N., & Tenenbaum, J. (2010). Theory acquisition as stochastic search. In *Proceedings of thirty-second annual meeting of the cognitive science society*.
Van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, *14*, 355–384.
Wellman, H. M., Cross, D., Bartsch, K., & Harris, P. L. (1986). Infant search and object permanence: A meta-analysis of the A-not-B error. *Monographs of the Society for Research in Child Development*, *51*, 1–67.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.