Received: 5 September 2019 | Accepted: 19 July 2020

DOI: 10.1111/infa.12364

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

# Integration of thought and action continued: Scale errors and categorization in toddlers

**Martha E. Arterberry**[1] | **Susan J. Hespos**[2] | **Cole A. Walsh**[1] | **Carolyn I. Daniels**[1]

[1]Colby College, Waterville, Maine

[2]Northwestern University, Evanston, Illinois

**Correspondence**
Martha E. Arterberry, Department of Psychology, Colby College, 5553 Mayflower Hill, Waterville, ME 04901, USA.
Email: Martha.Arterberry@colby.edu

**Funding information**
National Science Foundation, Grant/Award Number: BCS-1729720

**Abstract**

To further explore the effect of weighted arms on toddler's performance in problem solving (Arterberry et al., 2018, *Infancy*, 23(2), 173), the present study explored scale errors and categorization, two instances where infants appear to show more advanced knowledge than toddlers. Experiment 1 ($N = 67$) used a novel task for inducing scale errors among 24- to 29-month-olds. Results replicated rates of scale errors found in previous research that used different tasks. Experiment 2 used sequential touching ($N = 31$) and sorting measures ($N = 23$) to test categorization in 24-month-old children. In both measures, children showed categorization at the basic level when there was high contrast between the exemplars, but not at a basic level with low contrast or a subordinate level. In Experiments 1 and 2, half the participants were tested while wearing weighted wristbands. Weighting the arms did not affect error rates, in contrast to previous research showing that weights improved performance in search tasks. The findings are discussed in light of children's difficulty in integrating perception, cognition, and action.

## 1 | INTRODUCTION

Developmental change is often portrayed as an upward trajectory where the abilities of the child are elaborated and refined over time in continuous improvement (Siegler et al., 2017) or a progression of stages (Piaget, 1954). However, there are several instances where young infants appear to have knowledge that older children fail to use. For example, infants in the first months of life know that hidden objects continue to exist (e.g., Baillargeon, 1987; Baillargeon, Spelke, & Wasserman, 1985). Yet toddlers fail to search for objects in the correct location (Keen, 2003; Piaget, 1954; Rivière & Lécuyer, 2008). At 2 months of age, infants expect a ball to come to rest against a wall and look significantly longer at an event where a ball appears to pass through a solid barrier (Spelke, Breinlinger, Macomber, & Jacobson, 1992). Yet, toddlers fail to use this information to locate a ball that has stopped against a barrier (Arterberry, Hespos, & Herth, 2018; Berthier, DeBlois, Poirer, Novak, & Clifton, 2000). Twelve-month-old infants expect that a protrusion under a cloth indicates the location of a hidden object (Baillargeon, 1995), yet 2-year-old children searching for a ball under one of three cloths, fail to look under the cloth that has a ball-shaped protrusion (Arterberry et al., 2018; Rivière & Falaise, 2011; Rivière & Lécuyer, 2003, 2008).

The value in looking at instances where infants appear to have knowledge that older children fail to use lies in highlighting distinctions of competence and performance. In infant looking-time studies like those described above, there is considerable scaffolding for the infant. There are no instructions to follow; the requirement is merely for the infant to attend to events on a puppet stage and to look away when they lose interest. Infants have a natural tendency to look longer at events that they find novel or surprising, so they look longer at events that violate their expectations—like balls passing through walls. There is no doubt that toddlers also find balls passing through walls surprising (Hood, Cole-Davies, & Dias, 2003) so it is not a question of competence. The difference in performance could be because toddlers are required to use language, follow directions, and execute actions to solve the cognitive tasks while engaging with another person. Integrating social, cognitive, and motoric skills is a sophisticated ability that we take for granted as adults, and it appears that toddlers are still learning to coordinate these skills into their repertoire of responses. Instances were toddlers fail to use knowledge they possess provides an opportunity to investigate the ability to integrate information.

The starting point for this research was an interesting series of studies conducted by Rivière and Lécuyer (2008) that probe the difficulties toddlers have in search tasks. They demonstrated that 29-month-olds typically make an error during a search task involving invisible displacement. However, performance improved significantly when children wore weighted wristbands while doing the task. To investigate this phenomenon further, we tested 24-month-old children in an identical search task (Arterberry et al., 2018). To test how far the wristband phenomenon generalized, we also tested the same children in a second search task where they needed to find a ball that had rolled behind one of four doors (similar to Berthier et al., 2000). Half the children wore weighted wristbands, and the rest wore unweighted wristbands. The results showed that children in the no-weight condition replicated previous findings of poor performance on both search tasks. Unlike 29-month-olds, the 24-month-olds in the weighted condition did not immediately show improvement on the search tasks. However, after an initial search attempt, children wearing weights performed significantly better than chance. Together, these studies suggest that weighed wrists appear to alter the motor planning system. In this sense, the improved performance is due to changing the coupling between an automated bottom-up response to a top-down response, like representing the continued existence of a hidden object.

In this paper, we investigate whether the benefit of weighted wrists would generalize to two new tasks—scale errors and categorization. We chose both tasks because they show that toddlers perform as if they do not encode size (in the case of scale errors) or categorize at the subordinate level (in the case of sequential touching) in contrast to young infants. Using simpler tasks, young infants appear to encode the height/size of objects and to respond as if they can categorize at the subordinate level. For example, Hespos and Baillargeon's (2001, 2006) research on 7-month-old infants demonstrates that infants encode the height of an object that is presented next to a container. The comparison of the objects and container allows infants to determine how much of the object will be hidden when the object is placed inside the container. Consequently, infants look significantly longer to an event where a tall object is hidden inside a short container. In contrast, when toddlers open the door on a small toy car, they seem surprised and frustrated when they cannot put their body inside the car (DeLoache, Uttal, & Rosengren, 2004). A similar competence–performance distinction is found in categorization. Seven-month-old infants are able to categorize different breeds of cats (a subordinate-level distinction; Quinn, 2004; Quinn, Doran, Reiss, & Hoffman, 2010). Yet, toddlers in a sequential touching task have difficulty with subordinate-level categorization (e.g., two different species of sharks) at 30 months of age (Bornstein & Arterberry, 2010).

## 1.1 | Scale errors

As early as 13 months of age, children make surprising mistakes by trying to put their bodies into small-scale toys, such as cars, chairs, and slides (e.g., DeLoache et al., 2004; DeLoache, LoBue, Vanderborght, & Chiong, 2013; see Figure 1). There are three types of scale errors. Body scale errors occur when children interact with objects in relation to their own body. Object and tool scale errors occur when children interact with two or more objects, such as putting a doll to sleep in a bed that is too small or selecting a tool that is too small to achieve a goal (Casler, Eshleman, Greene, & Terziyan, 2011; Ware, Uttal, Wetter, & DeLoache, 2006). Body scale errors tend to emerge first and are followed by object/tool scale errors (Jiang & Rosengren, 2018). Together, these findings show that children make scale errors in every-day contexts, such as their home or early education center, and in structured laboratory tasks (e.g., Brownell, Zerwas, & Ramani, 2007; Rosengren, Gutierrez, Anderson, & Schein, 2009; Rosengren, Schein, & Gutierrez, 2010).

Scale errors do not happen every time children interact with small objects, nor does every child make scale errors. Across a 6-month period from 12 to 26 months of age, 97% of parents reported that their children made at least one body scale error, with a total number of 79 body scale errors and 19 object scale errors found (Rosengren et al., 2009). In an online study, 18% of parents reported an

**FIGURE 1** Child making a scale error by trying to put a toy hat on her head

error that met the criteria of a scale error, and the errors peaked at 20–24 months of age (Ware, Uttal, & DeLoache, 2010). There appears to be a general decline in the number of body scale errors that occur from 18 to 30 months of age (Grzyb, Cangelosi, Cattani, & Floccia, 2019). In contrast, object scale errors occurred at a higher rate than body scale errors, and they were more common among older children (Casler et al., 2011; Ware et al., 2006). For the purpose of our study, we focused on body scale errors given our interest in studying spontaneous scale errors that occurred in the absence of training (e.g., Casler et al., 2011). With respect to our weighted-wrist condition, we selected a task that allowed the child to sit at a table manually engage with objects. Manual scale errors were important for our goals, because we wanted the actions to happen on the limbs that were wearing the weights to maximize the potential for a measurable effect.

One explanation for why children make scale errors involves an error in their action planning (Glover, 2004). Children see an object, such as a car, and they plan an action, such as sit in it to go for a ride. Then, they execute the action. If the size is not appropriate for the action, children demonstrate a scale error. Several researchers proposed that the reason children execute the action, despite the size of the object not allowing for completing the action, is because they cannot inhibit the plan (DeLoache et al., 2004, 2013). What triggers the plan? One possibility is that children focus on object shape rather than size (Grzyb et al., 2019). Grzyb et al. (2019) showed that children who made scale errors also attended less to object size in a looking-time task compared to children who did not make scale errors. Moreover, focus on shape over size may be a product of the children's developing language abilities (Grzyb et al., 2019; Hunley & Hahn, 2016). Another possibility is that children focus on object function, or have a teleofunctional stance, and object function takes precedence over object size in guiding action (Casler et al., 2011). Across all explanations is a theme of attention to one feature of the object (shape or function) while ignoring another (size). This characteristic of scale errors makes it similar to our previous work where toddlers make mistakes retrieving hidden objects in a search task because they pay attention to one feature of task instead of integrating across perceptual, motor, and cognitive information (Arterberry et al., 2018). For example, in Rivière and Lécuyer's (2008) invisible displacement task, an object is hidden in one of three locations. Children may search at the last-seen location of the hand rather than at the location where there is a ball-shaped protrusion under the cloth.

Not all children commit scale errors on a regular basis or in a predictable way in their natural interactions with objects. To study scale errors in the laboratory, often studies primed children to increase the likelihood of them producing errors. For example, children first engaged with appropriately sized toys or tools, and then, the objects were replaced by smaller-sized ones (e.g., Casler et al., 2011; DeLoache et al., 2004, 2013). Our goal was to create a new variation of the scale error task that would maintain a high number of scale errors, but not take a long time.

## 1.2 | Categorization

Categorization is yet another example where toddlers fail to use information that infants possess. Infants within the first 6 months of life are accomplished categorizers. They categorize objects globally, including animals, vehicles, and furniture (Arterberry & Bornstein, 2001; Behl-Chadha, 1996), and they also categorize with surprising specificity, such as dogs, horses, cats, facial expressions, and motion patterns of animals and vehicles (e.g., Arterberry & Bornstein, 2001, 2002; Bornstein & Arterberry, 2003; Eimas & Quinn, 1994). Seven-month-old infants even categorize different breeds of cats (Quinn, 2004; Quinn et al., 2010). In contrast, when older children are tested using the sequential touching procedure, the developmental trajectory appears to emerge on a much later time course.

The sequential touching procedure allows children to play with eight replica objects from two different categories, and the dependent measure is the number of within-category touches they perform. In a study charting the development of categorization, Bornstein and Arterberry (2010) showed that children at 12, 18, and 24 months categorized superordinate sets (animals vs. vehicles) and basic-level high-contrast sets, such as frogs versus cows or helicopters versus pickup trucks. It was not until 30 months of age that children categorized basic-level low-contrast sets, such as dogs versus horses or panel trucks versus sport utility vehicles, and no age group categorized at the subordinate level (Mako vs. Hammerhead sharks or convertible vs. hardtop sports cars).

Similar to the explanations for why children make scale errors, the staggered progression of categorization ability has been attributed to attention (Bornstein & Arterberry, 2010). For example, children show better categorization with observable properties than nonobservable properties, such as shape versus squishiness, unless the nonobservable property is made salient (e.g., Ellis & Oakes, 2006). In order to categorize sets with lower levels of contrast, children need to align the different features of the two categories, such as noticing that dogs have shorter legs and smaller bodies than horses. In other words, children must make finer discriminations (Gibson, 1969; Keil, 1989; Mandler, 1992). Oakes and Madole (2003) argued that older children recognize more and different shared attributes of objects and therefore shift their attention among attributes with greater ease. Even adults show increasingly refined categorizations based on expertise (Tanaka & Taylor, 1991). Thus, successful categorization involves attending to the right information and using that information to plan action, such as reaching and grasping in the sequential touching context. For these reasons, categorization through sequential touching is a good candidate for testing whether weighted wristbands will improve performance.

## 1.3 | Integrating information

We propose that toddlers' performance compared to infants is based on a failure to *integrate* relevant information rather than failing to attend to it or using information to inhibit an action. Support for this explanation comes from search tasks that 2-year-olds typically fail despite having the perceptual skills to achieve them (Arterberry et al., 2018). Take, for example, the door task used by Berthier et al. (2000). In this task, children viewed a ball rolling down a ramp behind an occluding panel and stopping at a barrier (the barrier was visible above the panel). Twenty-four-month-olds performed at chance levels when asked to choose the correct door (out of four) to retrieve a ball. They continued to make errors in simplified versions of the task including: When a transparent panel was used, when the number of doors was reduced to only two, when the children could see the full movement and resting location of the ball before it was occluded, when the children were provided with landmarks, or when the children merely pointed to the location of the ball (Butler, Berthier, & Clifton, 2002; Mangalindan & Schmuckler, 2011; Mash, Keen, & Berthier, 2003; Mash, Novak, Berthier, & Keen, 2006; Shutts, Keen, & Spelke, 2006). However, when 24-month-old children were tested wearing weighted wristbands, they met with greater success than children wearing unweighted wristbands (Arterberry et al., 2018).

Similarly, weighted arms helped children at 24 and 29 months find a hidden object in a three-location invisible displacement task. In this task, developed by Rivière and Lécuyer (2008), a small object was concealed in the experimenter's hand. The experimenter moved their hand under three identical cloths (A, B, and C) and after passing under cloth C revealed that their hand was empty. The object was always left at location B, and there was a visible bump in the cloth indicating the location of the hidden object (Rivière & Falaise, 2011; Rivière & Lécuyer, 2003, 2008). Even though the visible bump was a clue to the correct search location, this task was difficult for 24- and 29-month-olds wearing unweighted wristbands—the most common error was to search in location C (hence, the name C-not-B task). In contrast, children with weighted wristbands met with success (Arterberry et al., 2018; Rivière & Lécuyer, 2008). Thus, weighted arms may have allowed children to integrate the relevant sources of information into their action plan.

*Exactly* how weights affect the way children respond is still unclear; however, latency analyses show that weights did not slow down the action (e.g., the reach to the correct door or location; Arterberry et al., 2018; Rivière & Lécuyer, 2008). Moreover, the effect of weights was not immediate in that 24-month-old children wearing weighted wristbands made errors on their first search but met with success on their second search (Arterberry et al., 2018). Exploring whether the advantage of weighted wrists extends to other tasks besides search tasks allowed us to understand more fully the boundary conditions of perception–action and representation-guided responding. We return to these issues in the discussion.

### 1.4 | Current Study

In the present study, we continued to explore the challenges toddlers face in integrating information by studying scale errors (Experiment 1) and categorization (Experiment 2). There were four goals: (a) conceptually replicate scale errors in 24- to 29-month-old children using a novel procedure; (b) replicate animal and vehicle categorization performance of 24-month-olds; (c) test whether weighted wristbands would reduce scale error frequency; and (d) increase categorization performance.

## 2 | EXPERIMENT 1

In Experiment 1, we presented 24- to 29-month-old children with a new scale error task. In order to test for the effect of weighted wristbands on scale errors, we needed a task in which an arm movement was the dominant response (as opposed to sliding down a slide, sitting on a chair, or fitting one's body into a vehicle; Deloache et al., 2004). Thus, we asked 24- to 29-month-old children to engage with Mr. Potato Head©, a toy that involves placing body parts and accessories, like glasses, hat, and shoes, on a potato body. We predicted that children of this age would show scale errors in this task by trying to put the accessories on their own body.

To test for the effect of weights on scale errors, half of the children played with Mr. Potato Head wearing weighted wristbands and half wore unweighted wristbands. If weighted arms affected the production of scale errors, we predicted that children wearing the weighted wristbands would show fewer scale errors than children wearing unweighted wristbands. The age range of 24–29 months encompassed the age range during which scale errors have been previously documented (e.g., Grzyb et al., 2019) and the age range used by others to test the effect of weighted arms on search tasks (Arterberry et al., 2018; Rivière & Lécuyer, 2008).

### 2.1 | Method

#### 2.1.1 | Participants

Sixty-seven children between the ages of 24 and 29 months (*M* age = 25.77 months, *SD* = 2.29; 41 males) participated. Children from both rural and urban communities were recruited via email, phone, and announcements through local early education programs. The sample was 12% Hispanic and 88% non-Hispanic. The racial diversity was 1% African American, 3% Asian, 3% mixed race, and 93% White. Ninety-four percent had a parent who completed a college degree or higher.

Both Experiments 1 and 2 were conducted according to the guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures involving human subjects were approved by the Institutional Review Boards of Colby College and Northwestern University. Families received $20 per child as compensation for their time. A statistical power analysis was performed based on Arterberry et al. (2018). Their reported effect size was Cohen's $d = 1.13$. With an alpha = .05 and power = 0.95, the projected sample size required is approximately $N = 44$ (using GPower 3.1.9). Consequently, our sample sizes of $N = 67$ (Experiment 1) and $N = 54$ (Experiment 2) are adequate.

#### 2.1.2 | Materials

*Wristbands*

Two weighted wristbands, one for each arm, were created by inserting two C-sized batteries into a black sports headband and sewing the ends together. Each wristband measured 23 cm in circumference, was 5 cm, and weighed 136 grams. Based on an average body weight of children aged 24 months (11,850.1 g), each band comprised 1.15% of the child's body weight. The amount of weight was comparable to the 200 g weights used with 29-month-olds by Rivière and Lécuyer (2008). No-weight wristbands were created by inserting two fabric rolls the same size and density of the batteries into a black sports headband and sewing the ends together. The resulting wristbands were similar in bulk to the weighted wristbands but only weighed 15 g each. All children wore a long-sleeved garment to hide the wristbands from view and reduce distraction.

*Scale error task*

The scale error task used a Mr. Potato Head toy (Playskool/Hasbro, 02650/02579; Figure 2). Overall dimensions were 20 cm high × 12 cm deep × 22 cm wide (with arms extended). Pieces included two pairs of shoes, two hats, two sets of glasses, eyes, a tongue, two ears, two arms, a smile, a nose, and

**FIGURE 2** Mr. Potato Head toy pictured assembled along with extra pieces

a mustache, and the potato-shaped body. The toy was presented disassembled in a small red basket measuring 19.5 cm long × 14 cm wide × 12.5 cm high.

*Videorecording*

A digital camera provided a frontal view of the child. An iMac captured the video and audio using PhotoBooth (©Apple). The recording showed children's upper body including the head and direction of gaze, hands, and objects on the table.

### 2.1.3 | Procedure

While parents provided informed consent and demographic information, the experimenter engaged the children with some toys. When ready, the experimenter asked the children to wear wristbands, randomly assigned to weight or no-weight conditions, and a long-sleeved garment to cover the wristbands. Parents wore a second set of unweighted wristbands in order to encourage the children to comply. Children who refused to wear the wristbands were assigned to the no-weight condition (three were intended for weight condition and three were intended for the no-weight condition). If children were wearing their own long-sleeved garment, parents hid the wristband under the sleeves. Next, the experimenter showed the children and their parents to the testing room. Children completed three tasks, in a predetermined random order; we only report on the scale error task here.

To begin the parent and children sat on one side of a small table and the experimenter sat on the opposite side. The experimenter asked parents to show/help their children put Mr. Potato Head together. Parents were encouraged to prompt their children, for example, "What should we do with these?" to help the children engage with the toy; however, they were asked not to instruct their children regarding what to do with the duplicate pieces. Next, the basket containing the Mr. Potato Head pieces was inverted such that the pieces were spread out on the table within arms' reach of the children. The children played with the toy until they showed signs of boredom or became distracted. To end, they were asked if they would like to help clean up, and they put all the pieces back into the basket.

### 2.1.4 | Coding

Two research assistants who were naive to the weight condition independently viewed the video sessions and coded (a) the total time the children engaged with the pieces, (b) the number of times children touched a piece to their body, and (c) the number of scale errors made with one of the duplicate pieces (hat, shoes, glasses). To obtain global measures of how the children interacted with the toy, along with their parent, the assistants also coded (a) the number of pieces successfully placed on the toy with and without parental assistance and (b) the number of times a piece was removed with and without parental assistance. We used the criteria from DeLoache et al. (2004) to determine a scale error. The action had to be the same action one would perform with the object's larger counterpart, and each instance had to be a serious attempt, such as persisting when failing or attending to the object. We restricted our investigation to actions involving only the duplicate accessories, namely the eyeglasses, the hat, and the shoes. Some children touched one or more body pieces to their own body, such as the teeth to their teeth; however, because it was unclear whether these were scale errors or an attempt to match a piece to their own body part, they were not counted as scale errors. Instead, these actions were labeled "body touches." Agreement between the two coders was 97%. The two coders resolved disagreements by reviewing the videos together.

## 2.2 | Results

Preliminary analyses tested for differences between boys and girls. As none were found, the following analyses were conducted collapsing across sex.

### 2.2.1 | Inducing scale errors

The children engaged with the toy and its parts for an average of 5.89 ($SD = 2.13$) minutes. During that time, they placed an average of 2.83 ($SD = 2.33$) pieces with assistance and 5.86 ($SD = 3.41$) pieces without assistance. In addition, children touched a piece to their body 86 times. Sixty-two (72.10%) of all actions were clearly scale errors: The children tried to put the hat on the top of their head ($N = 33$ times) or the glasses on their face ($N = 29$ times). No scale errors were made with the shoes. Of the 67 children tested, 30 children did not make any scale errors. Of the 37 who made at least one error, 13 children made only one error, six children made two errors, four children made three errors, four children made four errors, two children made five errors, and one child made seven errors.

Overall, there were 1.28 body touches per child and 1.04 scale errors per child. The amount of time children engaged with the toys did not differentiate children who did ($M = 6.68$, $SD = 2.59$) and did not ($M = 6.20$, $SD = 2.66$) make one or more scale errors, $t(64) = 0.74$, $p = .460$. The number of different or unique pieces children played with did not differentiate children who did ($M = 4.67$, $SD = 2.63$) and who did not ($M = 5.46$, $SD = 2.63$) make one or more scale errors, $t(23) = 0.82$, $p = .422$.

### 2.2.2 | The effect of weight

The amount of time that children engaged with the toys did not differentiate between the weight conditions (weight: $M = 6.75$, $SD = 3.09$; no-weight: $M = 6.13$, $SD = 2.15$), $t(64) = 0.97$, $p = .336$. To test for the effects of weighted arms on children's commission of scale errors, an independent samples *t* test revealed no difference in the number of scale errors across the weight ($M = 5.72$; $SD = 2.21$) and no-weight ($M = 6.03$; $SD = 2.09$) conditions, $t(65) = 0.05$, $p = .958$.

To explore further the effect of weight on children's likelihood to make a scale error, a chi-square analysis was conducted (see Table 1). Children were categorized as having made no scale errors or at least one scale error. The frequency of scale errors was not dependent on weight condition, $X^2(1) = 0.76$, $p = .383$ and $X^2(1) = 0.50$, $p = .479$, respectively.

## 2.3 | Discussion

Our first finding showed that Mr. Potato Head reliably induced scale errors in a laboratory setting. In our sample, 50% of children made a scale error with one of Mr. Potato Head's duplicate pieces, with

**TABLE 1** Frequency of children who made scale errors as a function of weight condition

| | No-weight | Weight | Total |
|---|---|---|---|
| No scale errors | 19 | 18 | 37 |
| One or more scale errors | 18 | 12 | 30 |
| Total | 37 | 30 | 67 |

an average of 1.04 scale errors per child. This rate was comparable to previous studies on children's scale errors conducted in laboratory settings. For example, in Deloache et al. (2004), 46% of the children made at least one scale errors, averaging 0.74 per child, and in Deloache et al. (2013), 50% of children made at least one scale error, averaging 0.84 scale errors per child. Our rate of scale errors was higher than that reported in Gryzb et al. (2018); in their study, 38% of children committed at least one scale error, averaging of 0.83 scale errors per child. The scale error rate was smaller than that reported in Casler et al. (2011), where 98% of children committed a scale error, averaging of 2.6 scale errors per child, likely due to the fact that Casler et al. studied object scale errors, and they provided children with training regarding the use or function of the objects. Unlike in previous studies, children in the present study did not first have the opportunity to engage with appropriately sized objects (e.g., Brownell et al., 2007; Casler et al., 2011; DeLoache et al., 2004; Grzyb et al., 2019; Ware et al., 2006). Thus, we were able to capture the rate of spontaneous scale error occurrence in a semi-structured laboratory task while examining the effects of weighted arms.

Our second finding was that weighted wristbands did not affect children's rate of scale errors. We hypothesized that weighted wristbands would help children integrate perceptual and cognitive information with their action. In two search tasks, C-not-B and the Door Task, weighted wristbands helped children commit fewer errors (Arterberry et al., 2018). The improvement from weighted wristbands did not generalize to the scale error task. We return to this point in the General Discussion.

## 3 | EXPERIMENT 2

Experiment 2 tested the effect of weights on children's categorization. Studies using infant looking time typically find sophisticated categorization skills in very young infants (e.g., Arterberry & Bornstein, 2001, 2002; Quinn et al., 2010). In contrast, when children are tested for categorization with tasks that encourage interactions with objects, such as the sequential touching procedure (Mandler, Fivush, & Resnick, 1987), children's categorization skills appear protracted, such that they do not categorize subordinate sets even at 30 months of age (Bornstein & Arterberry, 2010).

We tested 24-month-old children's categorization in a sequential touching task and in a sorting task. In the sequential touching task, children freely played with a collection of objects from two categories (e.g., four animals and four vehicles), and their touching patterns were recorded. The rationale is that if children recognize a categorical distinction among the objects, then they should touch those within a category in succession more often than would be expected by chance. We tested children's categorization of animals or vehicles three ways (see Figure 3). Categorization at the basic level was tested two ways: One set had a high level of contrast (e.g., frogs and cows or helicopters and pickup trucks), and the other set had a low level of contrast (e.g., dogs and horses or SUVs and semi-tractor trailers). Categorization at the subordinate level had one set (e.g., Mako and Hammerhead sharks or hardtop sports cars and convertibles).

The first goal of this experiment was to replicate Bornstein and Arterberry (2010)'s results. Bornstein and Arterberry (2010) found that 24-month-olds categorized objects at the basic level when high contrast was present, but not when the basic-level sets had lower levels of contrast or subordinate-level sets. The second goal was to test the effects of weights on categorization performance. If weights on the arms facilitated categorization, then children wearing weighted wristbands should show improved categorization of basic-level, low-contrast and possibly subordinate sets.

Alternatively, it is possible that weights may not affect categorization performance in the sequential touching procedure because of the open-ended nature of the task. In this procedure, children freely play with the objects, whereas in previous search tasks where weights were effective (e.g., Door Task

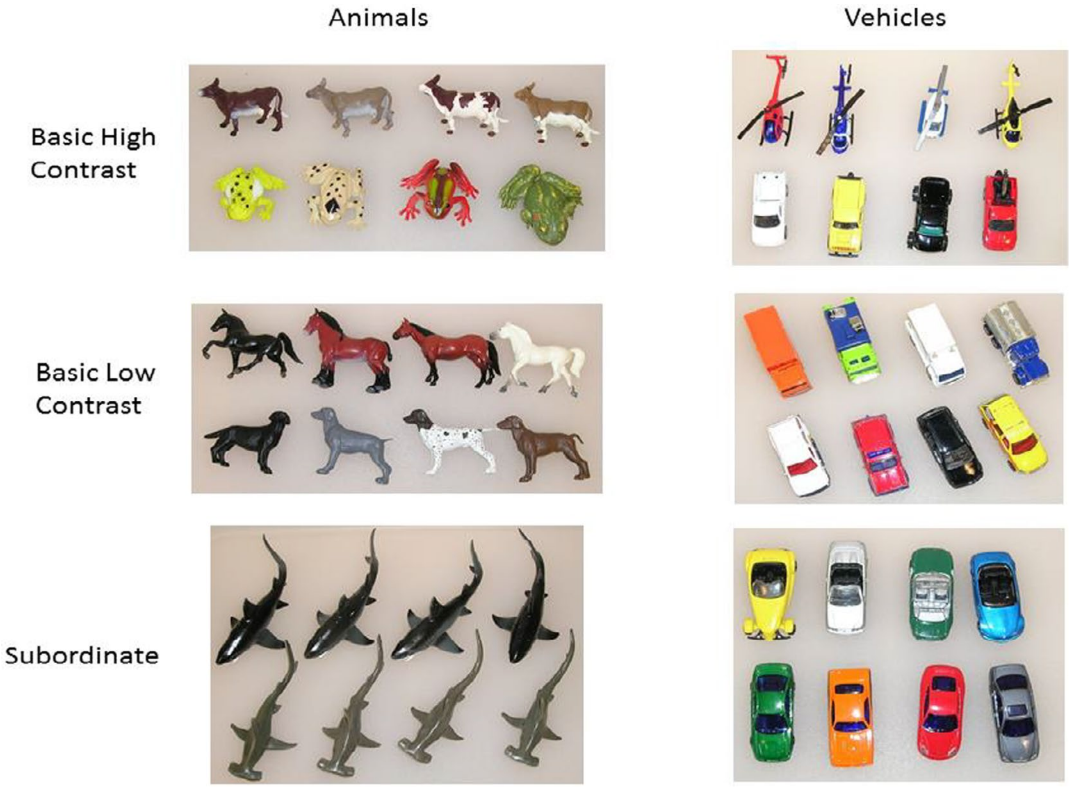


**FIGURE 3** Animal and vehicle stimuli used in Experiment 2

and C-not-B), the experimenter directed the children to find a hidden object. To investigate the role of explicit instructions and feedback, we tested a second group of 24-month-olds' categorization in a sorting task. Children again engaged in free play with eight objects, four from two different categories. After 2 min, children helped to "clean up." The experimenter presented two bins, and the experimenter instructed the children to put one category of objects in one bin and the other category of objects in the other bin. Thus, sorting was the goal and the feedback consisted of children seeing whether the objects they placed in the bin matched the one(s) already there.

In sum, Experiment 2 allowed for the replication of Bornstein and Arterberry (2010) with a sequential touching task and a conceptual replication using a sorting task. Additionally, we tested whether weighted arms improved categorization performance.

## 3.1 | Method

### 3.1.1 | Participants

Fifty-four 24-month-old children (*M* age = 23.61, *SD* = 0.60; 37 males) participated. Families were recruited through the same methods as Experiment 1. The sample was 14% Hispanic and 86% non-Hispanic. Racially, 2% were African American, 7% were mixed race, and 92% were white. Ninety-four percent had at least one parent who completed a college degree or higher. Thirty-one children completed the sequential touching task (14 in the weight condition and 17 in the no-weight condition), and 23 completed the sorting task (12 in the weight condition and 11 in the no weight condition). Four additional children began testing but their data are not included: Two in the sorting task due to lost video (1) and inability to sort any blocks even with experimenter feedback (1) and two in the sequential touching procedure due to not engaging with the objects for two minutes.

### 3.1.2 | Design

We tested 31 children in the sequential touching task, with approximately half in the weight and half in the no-weight condition. We also developed the sorting task and tested an additional 23 children, with approximately half in the weight and half in the no weight condition. In both tasks, children received sets representing basic-level high-contrast, basic-level low-contrast, and subordinate comparisons of either animal or vehicle categories across three trials. The resulting mixed 2 (task) × 2 (weight condition) × 3 (level) design allowed us to test for the between-subjects effects of weights across task and the within-subjects factor of category level.

### 3.1.3 | Materials

*Wristbands*

The wristbands were the same as those used in Experiment 1.

*Animal and vehicle stimuli*

The animal and vehicle stimuli were the same as those used by Bornstein and Arterberry (2010). Small naturalistic 3-dimensional scale models created 6 sets of stimuli (see Figure 3). The replicas averaged 9.15 × 4.93 × 5.74 cm. The basic-level high-contrast animal set contained four frogs and four cows. The basic-level, high-contrast vehicle set contained four helicopters and four pickup trucks. The basic-level, low-contrast animal set contained four dogs and horses. The basic-level, low-contrast vehicle set contained four sport utility vehicles (SUVs) and four tractor trailers. The subordinate animal set contained four Hammerhead sharks and four Mako sharks. The subordinate vehicle set contained four hardtop sports cars and four convertible sports cars. The objects were arranged on a 38.3 × 51.1 cm tray.

*Sorting task*

In addition to the weights and animal and vehicle stimuli described above, we used two bins and 10 Mega® Bloks for the sorting task. The two bins measured 16.5 × 26 × 6 cm (Figure 4). Five of the Mega-Blocks were small, measuring 3 cm long × 3 cm wide × 5.5 cm high, and five of them were large, measuring 12.5 cm long × 3 cm wide × 4.5 cm high.

### 3.1.4 | Procedure

After obtaining parental consent, children were seated across the table from the experimenter in a child-sized chair. A camera positioned behind the experimenter provided a frontal view of the child, including their head, their hands, and the objects on the table. Children in the sequential touching task completed three tasks, in a predetermined random order; we only report on the sequential touching task here.

**FIGURE 4** Setup of objects and bins at the beginning of the "clean up" phase of the session

*Sequential touching task*

All children completed three sets, either animals or vehicles, randomly assigned. The order of presentation of the basic-level high-contrast, basic-level low-contrast, and subordinate sets was randomly determined. To begin each trial, the experimenter spread out the eight objects in the set (four from each category) randomly on a tray in front of the child. No more than two objects from the same category were in the same immediate location. She then slid the tray across the table within reach and prompted: "These are for you to play with." The children were allowed to play with the objects for two minutes without any instruction from the experimenter or their parents. If the children became distracted, the experimenter prompted them to continue playing by repeating: "These are for you to play with." After the two minutes or if the children did not touch a toy for at least 10 sconds (indicating a loss of interest), the experimenter took the objects away and repeated the same procedure with the next set. If an object fell or was out of the children's reach, the experimenter or parent put the object back within reach. Three children did not complete all three sets: Two completed two of the three sets and one completed only one set.

*Sorting task*

Children began the task with the large and small blocks. Children initially played with the objects until bored or for 2 min, following the same procedure as described above for the sequential touching task before being asked to help clean up. During the "clean up" phase, the experimenter placed two bins in front of the child, one to the left side and one to the right side, and placed one-sized block (e.g., small) in one bin and one-sized block (e.g., larger) in the other bin. With each placement, the experimenter said "put the x (e.g., small) ones here." If children made a misplacement (such as placing a small block into the large-block bin), the experimenter corrected the children by removing the object from the wrong bin and repeating which sized block goes in that bin. The next three sets (basic-level high-contrast, basic-level low-contrast, and subordinate) were presented one at a time. For each set, children played with the objects for up to two minutes, and then "cleaned up" the same way as described with the blocks except that the experimenter did not give feedback in the event of an incorrect placement. The order of the three sets was randomly determined as well as which objects (e.g., frogs vs. cows) were placed in the left and the right bins.

#### 3.1.5 | Coding

Two coders, blind to the hypotheses of the study or weight condition, coded the sessions to determine the order in which each child engaged with each object (sequential touching) or the order the objects were placed in the bins (sorting). After coding independently, the two coders reviewed their agreement (95%) and resolved any differences.

For the sequential touching task, coders recorded the number of touches made to each object. To be counted as a touch, the child had to be looking at the object as they touched or picked it up. A touch was also recorded if the child used one object to contact another (e.g., touch the cow with the frog), or if an object was returned to multiple times. It was also noted when a child held two objects at once and the order in which they picked them up. The number of touches was used to calculate mean run length (e.g., Mandler et al., 1987). Run length is the number of touches in a row to objects from the same category (e.g., red frog, yellow frog, green frog equals a run of 3). A run can range from 1 (if the child touches only one object from one category before touching an object from another category) to the total number of the child's touches (if the child touches only objects from one category). The mean of all run lengths (MRL) was then calculated.

For the sorting task, the coders determined how many items were correctly placed in the bins during the “clean up” phase. Maximum correct was six.

### 3.2 | Results

Preliminary analyses assessed the effects of sex on number of touches, mean run length, and number of items sorted correctly, but none were found. Thus, all analyses were collapsed across boys and girls. Analyses for outliers revealed two outliers (values were three standard deviations or more above the mean, after Bornstein & Arterberry, 2010) in the sequential touching task. One child had a mean run length of 12 for a basic-level, low-contrast set. Another child produced 50 touches to objects in a subordinate set. These children were excluded from the analyses of these variables.

#### 3.2.1 | Sequential touching task

Table 2 shows the mean number of touches and mean run length. Following Bornstein and Arterberry (2010), mean run length for each categorization set was compared to chance (1.75: see also Mandler et al., 1987). Mean run length for basic-level high-contrast sets was significantly greater than chance, $t(30) = 3.75$, $p = .001$, $d = .67$; however, mean run length for basic-level, low-contrast sets and subordinate sets was not, $t\text{s} < 1.68$, $p > .580$.

#### 3.2.2 | Sorting task

The block set was always presented first, and there were 10 blocks. The experimenter used two blocks to indicate which bin to place the large and small blocks. Thus, children were able to sort up to eight

**TABLE 2** Mean number of touches and mean run length (MRL) as a function of categorization level and weight condition

| | Basic high contrast | | Basic low contrast | | Subordinate | |
|---|---|---|---|---|---|---|
| | Touches | MRL | Touches | MRL | Touches | MRL |
| No-weight | | | | | | |
| *M* | 15.67 | 2.78 | 17.73 | 1.98 | 15.33 | 1.72 |
| *SD* | 5.90 | 1.55 | 7.36 | 0.92 | 7.07 | 0.60 |
| Weight | | | | | | |
| *M* | 17.08 | 2.57 | 13.83 | 2.11 | 14.42 | 1.96 |
| *SD* | 6.87 | 1.22 | 6.97 | 0.94 | 7.10 | 0.84 |
| Total | | | | | | |
| *M* | 16.30 | 2.68* | 16.00 | 2.04 | 14.93 | 1.82 |
| *SD* | 6.26 | 1.39 | 7.32 | 0.91 | 6.96 | 0.71 |

*$p < .05$; significantly greater than chance (1.75).

objects correctly. Mean correct sorting of the blocks was 6.07 ($SD = 1.83$), which was significantly greater than chance, $t(23) = 6.09$, $p < .001$, $d = 1.13$. Even though children were provided feedback about their placements, not all children were willing to adjust their sorting to match the experimenter's prompts.

Next, we compared children's sorting of the animal and vehicle sets to chance. Each set contained eight objects, and two (one from each category) were used to explain which objects go in which bin. Thus, children were able to sort up to six objects correctly. Table 3 shows the mean number of correct object placements as a function of category level and weight condition. Correct sorting of basic-level high-contrast sets was significantly greater than chance (3.00), $t(22) = 3.67$, $p = .001$, $d = .74$, but correct sorting of basic-level low-contrast sets and subordinate sets did not differ from chance, $t$s $< 1.31$, $p > .203$.

### 3.2.3 | Categorization across tasks, levels, and weight condition

Our primary interest was to test for the interaction of weight condition and task on children's categorization performance. Because each task relied on a different measure of behavior, we calculated $z$-scores from the mean run length and correct sorting measures. A 2 × 2 × 3 ANOVA with task (sequential touching, sorting) and weight condition (no-weight, weight) as between-subjects factors and level (basic-level high-contrast, basic-level low-contrast, and subordinate) as a within-subjects factor revealed no significant main effects or interactions, all $F$s $< 1.49$, $p$s $> .228$. Thus, performance did not differ between tasks or as a function of level or weight condition.

## 3.3 | Discussion

In both the sequential touching task and the sorting task, 24-month-old children showed evidence of categorizing basic-level high-contrast sets but not basic-level low-contrast or subordinate sets. Bornstein and Arterberry (2010), using the same animal and vehicle stimuli in a sequential touching procedure, showed that 24-month-old children readily categorized objects above chance at the basic

**TABLE 3** Mean number of correct object placements as a function of categorization level and weight condition

| | Basic level | | |
|---|---|---|---|
| | **High contrast** | **Low contrast** | **Subordinate** |
| No-weight | | | |
| *M* | 4.00 | 2.91 | 2.82 |
| *SD* | 1.61 | 0.30 | 0.60 |
| Weight | | | |
| *M* | 4.09 | 3.45 | 3.09 |
| *SD* | 1.22 | 1.04 | 1.30 |
| Total | | | |
| *M* | 4.04* | 3.22 | 2.95 |
| *SD* | 1.36 | 0.80 | 1.00 |

*$p < .05$; significantly greater than chance (3.00).

level when high contrast between the categories was present. We extended this finding to a sorting task. When the contrast between categories was more subtle, as at the basic-level low-contrast and subordinate levels, children's categorization performance was at chance. We replicate this finding directly using sequential touching and conceptually using sorting.

There was no difference in performance across children wearing weighted and unweighted wristbands. This finding is in contrast to other studies where weighted wristbands significantly improved children's performance on search tasks (Arterberry et al., 2018; Rivière & Falaise, 2011; Rivière & Lécuyer, 2003, 2008). One possible explanation could be the differences between the search tasks used by Arterberry et al. (2018) and the sequential touching procedure used in the present study. In the search tasks, there was a clear goal (e.g., “find the ball”). It was evident when toddlers did not achieve the goal (e.g., they did not find the ball). When children were incorrect, they were allowed to try again. In the sequential touching task, children were allowed to freely manipulate the objects and only told “these are for you to play with.” Even when children in the sorting task were provided with explicit instructions regarding which objects should go in which bins, there was still no difference in performance by children wearing weighted compared to unweighted wristbands.

## 4 | GENERAL DISCUSSION

In these experiments, we tested toddlers’ performance in two tasks. We selected tasks where infants appear to show more advanced knowledge of object height/size and categorization in habituation tasks than toddlers’ exhibit when they perform scale errors (Experiment 1) and sequential touching assessments of categorization (Experiment 2). These experiments are unique in using a novel procedure for inducing scale errors that replicated previous work capturing a high number of scale errors in a short period of time. In addition, we replicated 24-month-olds’ successful categorization of basic-level high-contrasts sets and difficulty in categorizing basic-level low-contrast sets and subordinate sets using both sequential touching and sorting tasks. Finally, we tested for an effect of weighted arms in reducing scale errors and facilitating categorization; however, we found no difference in performance between children who wore weighted wristbands versus those who did not.

The use of the Mr. Potato Head toy to induce scale errors allowed us to test spontaneous scale errors. Often laboratory-based studies first expose children to objects that fit—either their body or another object—and then change the size of the objects (e.g., Casler et al., 2011; Deloache et al., 2004). We showed that it is possible to study scale errors in the laboratory without having prior experience with identical appropriately sized objects by using a readily available toy. Moreover, the rate of scale errors was on par with or exceeded the rate found by others (e.g., Deloache et al., 2004; Gryzb et al., 2019; Rosengren et al., 2010) in our short 5-minute manually based task.

The findings from the categorization experiments replicated Bornstein and Arterberry (2010; see also Arterberry & Bornstein, 2012; Arterberry, Bornstein, & Blumenstyk, 2013). Children at 24 months categorized or sorted basic-level, high-contrast sets but not basic-level, low-contrast or subordinate sets. This finding held across two tasks, sequential touching and sorting, that used different measures to assess categorization.

The surprising finding was that weights did not affect performance. There was no difference in the incidence of scale errors between children wearing weighted wristbands and those wearing unweighted wristbands. Similarly, children were equally successful categorizing basic-level, high-contrast sets and equally unsuccessful in categorizing basic-level low-contrast and subordinate sets regardless of whether they wore weighted or unweighted wristbands. In previous studies using search tasks, children who wore weighted wristbands found the ball significantly more often than children who wore unweighted wristbands (Arterberry et al., 2018). One possible explanation for the lack of an effect in the weighted condition is that the weights were not heavy enough to impact performance. We think this interpretation is unlikely given the existing evidence from studies ranging from 24 to 29 months of age revealing that this same weight (1% of body weight) had a significant effect on performance across three different tasks (Arterberry et al., 2018; Rivière & Lécuyer, 2008).

Differences between the search tasks, where weights had a significant impact on performance (e.g., Arterberry et al., 2018), and the tasks used in the present study are that the search tasks had a clear goal and the opportunity for feedback. Failing to find a ball in a search task is an explicit failure in meeting a goal. Both the scale error task and the sequential touching task used to test categorization were open ended in that children were allowed to interact with the materials in any way they wished. Free play with objects is unstructured. In other words, the goal and success of meeting that goal is variable and up to the actor. Even when a goal was communicated by instructing children to sort objects into specific bins, children only sorted the basic-level high-contrast sets at a rate significantly greater than chance. In the sorting task, the experimenter communicated the goal but the feedback was subtle. Unless children noticed how their placements (e.g., a dog into one bin) matched the object already there (e.g., another dog in the same bin), children did not have direct evidence regarding whether they met the goal.

The picture that is emerging suggests that there is a dynamic interplay necessary to integrate social, cognitive, and motoric information. The solutions that toddler's choose in one task is multiply determined and may not generalize broadly. The null findings for weights help define the problem space where weighted wrists do not influence performance. Future research will investigate a finer-grained analysis of the influence of weighted wrists in the tasks where it does benefit performance with an eye toward understanding the mechanism underlying the effect of weights. For example, it is possible that weights aid in the planning of action such that arm movements are more coordinated or precise when wearing weights compared to no weights even though the overall latency to begin the movement is not affected by weights (Arterberry et al., 2018; Rivière & Lécuyer, 2008). In addition, monitoring eye movements to determine what aspects of the task children are attending to may provide insight into how weights may shift focus of attention from irrelevant to relevant aspects of the task.

In conclusion, we return to the competence–performance distinction that we raised in the introduction. We started with two seemingly contradictory developmental findings where infants appear to show more advance knowledge than toddlers. The difference between these two findings may stem from the scope of the generalization that is required in each task. The infant studies are highly structured and the toddler tasks are considerably more complex. Both scientific approaches are informative, and we can see the studies as spanning a range. At one end, are infant studies that are perfectly aligned exemplars presented in rapid succession—an ideal learning opportunity, although one that may not apply far beyond the initial stimuli. The infant studies provide an existence proof that the potential is evident early. At the other end, are toddler studies that present complex learning conditions in which the correct solution can be thwarted due to a variety of obstacles so success appears to emerge later. The toddler studies have value because they shed light on real-world learning conditions.

## ACKNOWLEDGMENTS

This work was supported by a Social Sciences Division grant from Colby College awarded to Martha Arterberry and a National Science Foundation grant (No. BCS-1729720) awarded to Susan Hespos. The authors declare no conflicts of interest with regard to the funding source for this study. We thank the members of the Infant Cognition Lab at Northwestern and Cognitive Development Lab at Colby College for their help with data collection and coding.

## ORCID

*Martha E. Arterberry* https://orcid.org/0000-0001-7089-4554

## REFERENCES

Arterberry, M. E., & Bornstein, M. H. (2001). Three-month-old infants' categorization of animals and vehicles based on static and dynamic attributes. *Journal of Experimental Child Psychology*, *80*, 333–346. https://doi.org/10.1006/jecp.2001.2637

Arterberry, M. E., & Bornstein, M. H. (2002). Infant perceptual and conceptual categorization: The roles of static and dynamic attributes. *Cognition*, *86*, 1–24. https://doi.org/10.1016/S0010-0277(02)00108-7

Arterberry, M. E., & Bornstein, M. H. (2012). Categorization of real and replica objects by 14- and 18-month-old infants. *Infant Behavior and Development*, *35*, 606–612. https://doi.org/10.1016/j.infbeh.2012.04.007

Arterberry, M. E., Bornstein, M. H., & Blumenstyk, J. B. (2013). Categorization of two-dimensional and three-dimensional stimuli by 18-month-old infants. *Infant Behavior and Development*, *36*, 786–795. https://doi.org/10.1016/j.infbeh.2013.09.008

Arterberry, M. E., Hespos, S. J., & Herth, R. A. (2018). Integration of thought and action: Arm weights facilitate search accuracy in 24-month-old children. *Infancy*, *23*(2), 173–193. https://doi.org/10.1111/infa.12208

Baillargeon, R. (1987). Object permanence in 3.5- and 4.5-month-old infants. *Developmental Psychology*, *23*, 655–664. https://doi.org/10.1037/0012-1649.23.5.655

Baillargeon, R. (1995). A model of physical reasoning in infancy. In C. Rovee-Collier, & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. *9*, pp. 305–371). Norwood, NJ: Ablex.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition*, *20*, 191–208. https://doi.org/10.1016/0010-0277(85)90008-3

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*, 105–141. https://doi.org/10.1016/0010-0277(96)00706-8

Berthier, N. E., DeBlois, S., Poirier, C. R., Novak, M. A., & Clifton, R. K. (2000). Where's the ball? Two and three-year-olds reason about unseen events. *Developmental Psychology*, *36*, 394–401. https://doi.org/10.1037/0012-1649.36.3.394

Bornstein, M. H., & Arterberry, M. E. (2003). Recognition, categorization, and apperception of the facial expression of smiling by 5-month-old infants. *Developmental Science*, *6*, 585–599.

Bornstein, M. H., & Arterberry, M. E. (2010). The development of object categorization in young children: Hierarchical inclusiveness, age, perceptual attribute and group versus individual analyses. *Developmental Psychology*, *46*, 350–365. https://doi.org/10.1037/a0018411

Brownell, C. A., Zerwas, S., & Ramani, G. B. (2007). "So big": The development of body self-'awareness in toddlers. *Child Development*, *78*(5), 1426–1440. https://doi.org/10.1111/j.1467-8624.2007.01075.x

Butler, S. C., Berthier, N. E., & Clifton, R. K. (2002). Two-year-olds' search strategies and visual tracking in a hidden displacement task. *Developmental Psychology*, *38*, 581–590. https://doi.org/10.1037/0012-1649.38.4.581

Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental Psychology*, *47*(3), 857–866. https://doi.org/10.1037/a0021174

DeLoache, J. S., LoBue, V., Vanderborght, M., & Chiong, C. (2013). On the validity and robustness of the scale error phenomenon in early childhood. *Infant Behavior and Development*, *36*(1), 63–70. https://doi.org/10.1016/j.infbeh.2012.10.007

DeLoache, J. S., Uttal, D. H., & Rosengren, K. S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science*, *304*(5673), 1027–1029. https://doi.org/10.1126/science.1093567

Eimas, P. D., & Quinn, P. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917. https://doi.org/10.2307/1131427

Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology*, *42*, 1000–1011. https://doi.org/10.1037/0012-1649.42.6.1000

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-Century-Crofts.

Glover, S. G. (2004). What causes scale errors in children? *Trends in Cognitive Sciences*, *8*, 440–442. https://doi.org/10.1016/j.tics.2004.08.002

Grzyb, B. J., Cangelosi, A., Cattani, A., & Floccia, C. (2019). Children's scale errors: A by-product of lexical development? *Developmental Science*, *22*, e12741. https://doi.org/10.1111/desc.12741

Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment: A surprising discrepancy. *Psychological Science*, *12*(2), 141–147.

Hespos, S. J., & Baillargeon, R. (2006). Decalage in infants' knowledge about occlusion and containment events: Converging evidence from action tasks. *Cognition*, *99*, B31–B41. https://doi.org/10.1016/j.cognition.2005.01.010

Hood, B., Cole-Davies, V., & Dias, M. (2003). Looking and search measures of object knowledge in preschool children. *Developmental Psychology*, *39*(1), 61–70. https://doi.org/10.1037/0012-1649.39.1.61

Hunley, S. B., & Hahn, E. R. (2016). Labels affect preschoolers' tool-based scale errors. *Journal of Experimental Child Psychology*, *151*, 40–50. https://doi.org/10.1016/j.jecp.2016.01.007

Jiang, M. J., & Rosengren, K. S. (2018). Action errors: A window into the early development of perception-action system. *Advances in Child Development*, *55*, 145–171.

Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, *12*, 79–83. https://doi.org/10.1111/1467-8721.01234

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*, 587–604. https://doi.org/10.1037/0033-295X.99.4.587

Mandler, J. M., Fivush, R., & Reznick, J. S. (1987). The development of contextual categories. *Cognitive Development*, *2*, 339–354. https://doi.org/10.1016/S0885-2014(87)80012-6

Mangalindan, D. M. J., & Schmuckler, M. A. (2011). What's in a cue? The role of cue orientation in object displacement tasks. *Infant Behavior and Development*, *34*, 407–416. https://doi.org/10.1016/j.infbeh.2011.03.002

Mash, C., Keen, R., & Berthier, N. E. (2003). Visual access and attention in two-year-olds' event reasoning and object search. *Infancy*, *4*, 371–388. https://doi.org/10.1207/S15327078IN0403_04

Mash, C., Novak, E., Berthier, N. C., & Keen, R. (2006). What do two-year-olds understand about hidden-object events? *Developmental Psychology*, *42*, 263–271. https://doi.org/10.1037/0012-1649.42.2.263

Oakes, L. M., & Madole, K. L. (2003). Principles of developmental change in infants' category formation. In D. H. Rakison, & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 132–158). New York, NY: Oxford University Press.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic Books.

Quinn, P. C. (2004). Development of subordinate-level categorization in 3- to 7-month-old infants. *Child Development*, *75*, 886–899. https://doi.org/10.1111/j.1467-8624.2004.00712.x

Quinn, P. C., Doran, M. M., Reiss, J., & Hoffman, J. E. (2010). Neural markers of subordinate-level categorization in 6- to 7-month-old infants. *Developmental Science*, *13*, 499–507. https://doi.org/10.1111/j.1467-7687.2009.00903.x

Rivière, J., & Falaise, A. (2011). What comes first? How selective attentional processes regular the activation of a motor routine in a manual search task. *Developmental Psychology*, *47*, 969–975. https://doi.org/10.1037/a0023767

Rivière, J., & Lécuyer, R. (2003). The C-not-B error: A comparative study. *Cognitive Development*, *18*, 285–297. https://doi.org/10.1016/S0885-2014(03)00003-0

Rivière, J., & Lécuyer, R. (2008). Effects of arm weight on C-not-B task performance: Implications for the motor inhibitory deficit account of search failures. *Journal of Experimental Child Psychology*, *100*, 1–16. https://doi.org/10.1016/j.jecp.2008.01.005

Rosengren, K. S., Gutiérrez, I. T., Anderson, K. N., & Schein, S. S. (2009). Parental reports of children's scale errors in everyday life. *Child Development*, *80*(6), 1586–1591. https://doi.org/10.1111/j.1467-8624.2009.01355.x

Rosengren, K. S., Schein, S. S., & Gutiérrez, I. T. (2010). Individual differences in children's production of scale errors. *Infant Behavior and Development*, *33*(3), 309–313. https://doi.org/10.1016/j.infbeh.2010.03.011

Shutts, K., Keen, R., & Spelke, E. S. (2006). Object boundaries influence toddlers' performance in a search task. *Developmental Science*, *9*, 97–107. https://doi.org/10.1111/j.1467-7687.2005.00466.x

Siegler, R. S., Eisenberg, N., Gershoff, E., Saffran, J. R., DeLoache, J., & Leaper, C. (2017). *How children develop* (5th ed.). New York, NY: Worth.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632. https://doi.org/10.1037/0033-295X.99.4.605

Tanaka, J. W., & Taylor, M. E. (1991). Categorization and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*, 457–482. https://doi.org/10.1016/0010-0285(91)90016-H

Ware, E. A., Uttal, D. H., & DeLoache, J. S. (2010). Everyday scale errors. *Developmental Science*, *13*(1), 28–36. https://doi.org/10.1111/j.1467-7687.2009.00853.x

Ware, E. A., Uttal, D. H., Wetter, E. K., & DeLoache, J. S. (2006). Young children make scale errors when playing with dolls. *Developmental Science*, *9*(1), 40–45. https://doi.org/10.1111/j.1467-7687.2005.00461.x

**How to cite this article:** Arterberry ME, Hespos SJ, Walsh CA, Daniels CI Integration of thought and action continued: Scale errors and categorization in toddlers. *Infancy*. 2020;00: 1–20. https://doi.org/10.1111/infa.12364