Received: 17 July 2023 | Revised: 6 February 2024 | Accepted: 7 February 2024

DOI: 10.1111/desc.13488

Developmental Science WILEY

RESEARCH ARTICLE

"Catastrophic" set size limits on infants' capacity to represent objects: A systematic review and Bayesian meta-analysis

Jinjing (Jenny) Wang[1] | Melissa M. Kibbe[2]

[1]Department of Psychology & Center for Cognitive Science, Rutgers University, New Brunswick, New Jersey, USA

[2]Department of Psychological & Brain Sciences, Boston University, Boston, Massachusetts, USA

Correspondence
Jinjing (Jenny) Wang, Department of Psychology & Center for Cognitive Science, Rutgers University, New Brunswick, New Jersey, USA.
Email: jinjing.jenny.wang@rutgers.edu

Melissa M. Kibbe, Department of Psychological & Brain Sciences, Boston University, Boston, Massachusetts, USA.
Email: kibbe@bu.edu

Abstract

Decades of research has revealed that humans can concurrently represent small quantities of three-dimensional objects as those objects move through space or into occlusion. For infants (but not older children or adults), this ability apparently comes with a significant limitation: when the number of occluded objects exceeds three, infants experience what has been characterized as a "catastrophic" set size limit, failing to represent even the approximate quantity of the hidden array. Infants' apparent catastrophic representational failures suggest a significant information processing limitation in the first years of life, and the evidence has been used as support for prominent theories of the development of object and numerical cognition. However, the evidence for catastrophic failure consists of individual small-n experiments that use null hypothesis significance testing to obtain null results (i.e., $p > 0.05$). Whether catastrophic representational failures are robust or reliable across studies, methods, and labs is not known. Here we report a systematic review and Bayesian meta-analysis to examine the strength of the evidence in favor of catastrophic representational failures in infancy. Our analysis of 22 experiments across 12 reports, with a combined total of $n = 367$ infants aged 10–20 months, revealed strong support for the evidence for catastrophic set size limits. A complementary analysis found moderate support for infants' success when representing fewer than four objects. We discuss the implications of our findings for theories of object and numerical cognitive development.

KEYWORDS
infancy, meta-analysis, object cognition, representation, working memory

Research Highlights

- Previous work has suggested that infants are unable to concurrently represent four or more objects—a "catastrophic" set size limit.
- We reviewed this work and conducted a Bayesian meta-analysis to examine the robustness of this limit across individual small-n experiments.
- We found strong support for the evidence for catastrophic set size limits, and moderate support for infants' success when representing fewer than four objects.

Jinjing (Jenny) Wang and Melissa M. Kibbe contributed equally to this work.

1 | INTRODUCTION

A fundamental question in cognitive science concerns the origins of our ability to represent objects that are no longer in view. Early in development, the ability to maintain persisting representations in the mind of nonvisible objects allows infants to navigate, learn about, and, eventually, to talk about the three-dimensional world in which they are embedded. These object representations may also provide the conceptual foundations for abstract symbolic systems of knowledge, such as knowledge about numbers (Carey, 2009; Carey & Barner, 2019; Feigenson et al., 2004).

Piaget established that infants can represent objects based on sensory data before they have the language abilities to talk about objects (Piaget, 1954). Investigations after Piaget have revealed that infants have robust and sophisticated object representations even in the absence of concurrent sensory input (Baillargeon et al., 1985; Bower, 1967; Feigenson & Carey, 2003, 2005; Káldy & Leslie, 2003, 2005; Kibbe, 2015; Kibbe & Leslie, 2011, 2016, 2019; Spelke et al., 1992), suggesting that infants can form representations of objects in their minds even without direct access to sensory data. For example, 5-month-old infants who observe an object that is occluded by a swinging drawbridge expect the object to stop the movement of the drawbridge (Baillargeon et al., 1985). Infants not only represent an object's continued existence when it is no longer in view (Baillargeon et al., 1985; Baillargeon, 1987; Kibbe & Leslie, 2011), but also can use objects' featural or categorical identities to individuate objects (Bonatti et al., 2002; Wilcox, 1999; Xu & Carey, 1996), can store such identifying properties in their representations of hidden objects (Káldy & Leslie, 2005; Kibbe & Leslie, 2019; Wilcox, 1999), and can use their representations of occluded objects to drive a range of behaviors and inferences, such as planning their own actions on objects (Shinskey & Munakata, 2005), understanding others' actions on objects (Applin & Kibbe, 2019), and choosing when to learn about objects (Stahl & Feigenson, 2015).

Over the last several decades, the *limitations* on infants' capacity to concurrently represent multiple objects have been a major focus of research, since such limitations can provide important insights into the earliest structures of the mind and the capacities from which the adult mind emerges. This work has revealed what has come to be regarded as a signature limit on infants' capacity to represent sets of occluded objects. When infants are tasked with tracking one, two, or three objects in a single location, infants appear able to do so quite robustly (Feigenson & Carey, 2003, 2005; Feigenson et al., 2002; van de Walle et al., 2000; Wynn, 1992). By contrast, when infants are tasked with tracking more than three objects in a single location, infants appear to be unable to keep track of even a *subset* of the objects that they were tasked with representing (Barner et al., 2007; Feigenson et al., 2002; Zosh & Feigenson, 2015; see also Kibbe & Stahl, 2023), a phenomenon which is often referred to as a "catastrophic" set size limit, or a catastrophic failure to remember multiple objects that are hidden from view (e.g., Feigenson & Carey, 2005; Zosh & Feigenson, 2009).

"Catastrophic" set size limits have been observed across multiple methods, and the existence of such failures plays a prominent role in a variety of formal theories of the development of number knowledge (Carey, 2009; Carey & Barner, 2019; Feigenson et al., 2004; Hyde, 2011; Mou, 2014), object cognition (e.g., Feigenson & Carey, 2003, 2005; Leslie et al., 1998), and working memory (Cowan, 2016; Kibbe, 2015; Zosh & Feigenson, 2009). However, it is difficult to gauge how robust or consistent these "catastrophic" set size limits are, given the extant literature. Catastrophic set size limits in infant experiments are evidenced by null results obtained with very small numbers of infants (typically fewer than 20 infants per experiment) analyzed using null hypothesis significance testing (i.e., a p-value greater than 0.05 is taken as evidence of a "failure", while a p-value less than 0.05 is taken as evidence of a success), and such null results are notoriously difficult to interpret, particularly with small numbers of participants (Button et al., 2013; Gigerenzer et al., 2004; Nickerson, 2000). Thus, despite its potential theoretical significance, the reliability of "catastrophic" set size limits—and whether these failures are indeed robust across studies, methods, and labs—is not known.

We examined the robustness of catastrophic set size limit effects in infancy using a preregistered systematic review and Bayesian meta-analysis. We had two primary goals. Our first goal was to gather all of the evidence from the infant literature for catastrophic set size limits, and summarize that evidence and its role in theory-building. To do so, we conducted a systematic review of the literature that has examined limitations on infants' capacity to concurrently represent multiple objects in their minds, which we summarize below, and then briefly discuss the role that these results have played in the formulation of prominent theories of number knowledge acquisition and memory development. Our second goal was to estimate the robustness of catastrophic set size limit effects in infancy. To do so, we conducted a Bayesian meta-analysis on the effect sizes observed in studies that found a signature of catastrophic set size limits in infancy. We conclude with a discussion of the implications of our findings for theories that rely on the robustness of catastrophic set size limits on object representations in infancy.

2 | REVIEW OF THE EMPIRICAL EVIDENCE

Feigenson et al. (2002) were the first to observe the phenomenon of catastrophic set size limits in infants. They used a "foraging"-style task that previously had been used to examine representational capacities in nonhuman primates (Hauser et al., 2000). In their task (illustrated in Figure 1), 10- and 12-month-old infants observed an experimenter distributing crackers into two opaque containers, which were placed at equal distances from the infant. They varied the total number of crackers hidden across experiments, but one container always contained more crackers than the other container. Once the crackers were distributed, infants were allowed to crawl toward one of the containers and retrieve the crackers inside. They found that when infants were given a choice between containers with one versus two crackers or two versus three crackers, infants crawled to the container with the greater number of crackers at rates significantly above chance, suggesting that infants could represent and distinguish one, two, or three hidden objects from each other. However, when any of the

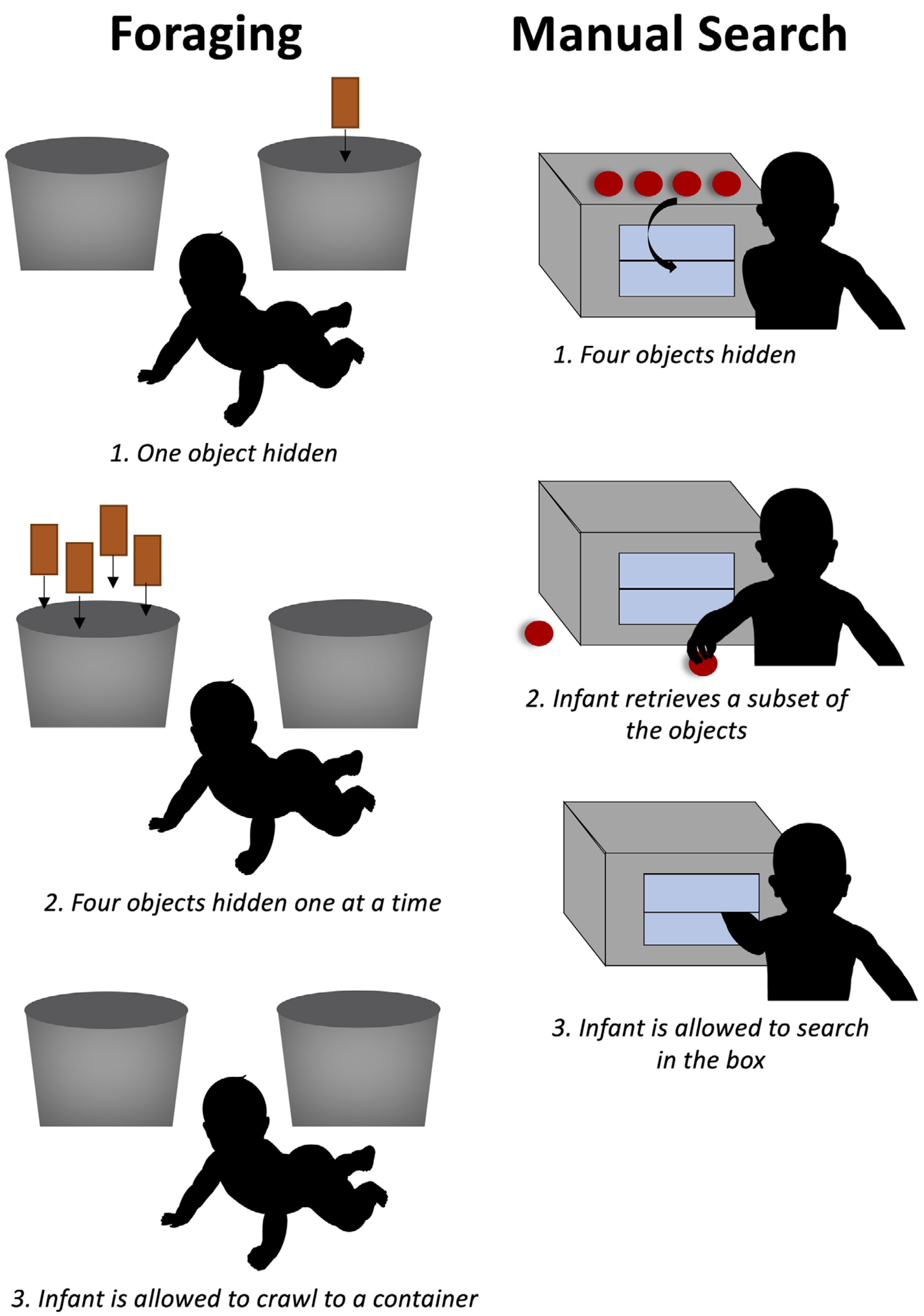

FIGURE 1 Illustrations of the foraging task and the manual search task. Here, we show examples of one possible trial for each study type. In the lab, parameters such as order and manner of object presentation, relative positions of the objects, and timing of object hiding/retrieval are carefully controlled and/or counterbalanced.

containers held more than three objects (e.g., two vs. four crackers, three vs. four crackers) infants crawled to the containers at roughly equal rates. Feigenson and Carey (2005) later found that this was true even when infants were tasked with tracking only *one* cracker in one location and four in the other—a distinction that should have been trivially easy for infants if they were able to track the approximate quantity of crackers (Cordes & Brannon, 2008). However, when one of the containers held four crackers and the other held none, infants successfully crawled to the container with four at rates above chance (Feigenson & Carey, 2005).

Together, these results were interpreted as infants experiencing a "catastrophic failure" in their ability to concurrently represent four or more objects—infants failed to store a subset of the objects, or an approximate estimate of the quantity of objects, but may retain that there was "something" left without knowing what or how much. Similar results were obtained by vanMarle (2013) with similar-aged infants. Similar results also have been obtained in rhesus macaques (Hauser et al., 2000), horses (Uller & Lewis, 2009), red-backed salamanders (Uller et al., 2003), and bees (Gross et al., 2009), suggesting that this signature limitation on object representational capacities may be phylogenetically ancient.

Catastrophic set size limits also have been observed in other tasks. Feigenson and Carey (2003, 2005) adapted the manual search task first used by Van de Walle et al. (2000). This method is illustrated in Figure 1. In their task, infants watched as an experimenter hid sets of one or more objects inside of an opaque box. Infants were allowed to reach into an opening at the front of the box and, without seeing what was inside the box, remove either all of the hidden objects ("expected empty" trials), or only a subset of the hidden objects ("more remaining" trials). To examine whether infants could keep track of how many objects were hidden, how many were removed, and therefore how many remained, they measured how long infants continued to search in the box following the removal of the objects. Feigenson and Carey (2003) found that infants searched longer in the box on "more remaining" trials compared to "expected empty" trials when sets of one, two, or three objects were hidden inside the box. However, infants searched roughly equally in the two trial types when four objects were hidden and two were retrieved compared to trials in which two objects were hidden and all were retrieved. Similar results were obtained by Barner et al. (2007), Feigenson and Carey (2005), Stahl and Feigenson (2014, 2018), Wang and Feigenson (2019, 2023), Stahl et al. (2023), and Zosh and Feigenson (2012).

Several studies also have found signatures of catastrophic set size limits in infants using passive-viewing tasks. Xu (2003) found that 6-month-old infants failed to discriminate two from four objects in a habituation/dishabituation study, but successfully discriminated larger sets of objects with the same ratio difference (four from eight objects) suggesting evidence for a divide between the way quantities smaller than four and larger than four are represented in infancy (see also Mack, 2006). Kibbe and Feigenson (2016) showed 13-month-old infants a set of four heterogeneous objects that was hidden behind an occluder, and the occluder was then removed to reveal either four objects or only three objects. Infants looked equally at these two outcomes, suggesting they failed to represent the quantity of the objects behind the occluder (see also Wang, 2023, for similar results using an online violation-of-expectation with 14- to 18-month-old infants). Similar results also were obtained with rhesus macaques in a violation-of-expectation looking time task (Hauser & Carey, 2003).

Catastrophic set size limits are harder to come by in humans beyond the infant years. Unlike infants, older children and adults who are tasked with attentionally tracking multiple objects show evidence of being able to track a subset of the objects when their attentional capacities are exceeded (Blankenship et al., 2020; Scholl & Pylyshyn, 1999; Trick et al., 2005; but see Feigenson, 2008, for evidence of a catastrophic limit on the number of *sets* of objects that adults can simultaneously track). And older children and adults who are tasked with storing representations of objects for brief intervals in working memory show evidence of remembering a subset of items when their working memory capacities are exceeded (Bays & Husain, 2008; Cheng & Kibbe, 2022, 2024; Luck & Vogel, 1997). These results suggest that signatures of catastrophic set size limits may be confined to infancy, and these severe limitations on human representational capacities ease with development. However, it is important to note that the methods used to assess representational capacities in infants tend to provide coarser evidence than the methods used to assess these capacities in older children and adults, since infant studies necessitate nonverbal methods and fewer trials.

2.1 | The significance of catastrophic set size limits for theories of development

The evidence for catastrophic set size limits in infancy has figured prominently in several leading theories of the structure of object representations, the development of numerical abilities, and the development of memory.

For example, one of the prominent theories of how humans represent object arrays and numerical quantities is the "two systems" theory (Feigenson et al., 2004; Hyde, 2011; Pylyshyn, 1994; Trick & Pylyshyn, 1994), which suggests that humans recruit categorically different representations when keeping track of a small number of objects versus a large array of objects. Behavioral signatures that support this theory include infants' apparent catastrophic failures when tracking a set of four or more individual objects, and adults' noise-free or precise performance when estimating the number of small arrays with four or fewer objects but noisy or ratio-dependent performance when estimating large arrays with more objects (Trick & Pylyshyn, 1994). The "two systems" theory has been challenged by the "unifying theory" (Gallistel, 2020; Gallistel & Gelman, 2000), which uses evidence that adults' performance on the small-quantity estimation task can become noisy and ratio-dependent with reduced exposure time to argue that the same representations are recruited to represent quantity regardless of set size (Cheyette & Piantadosi, 2020). The "unifying theory," therefore, does not predict infants' "catastrophic" failure when tasked with representing four or more objects, since under the unifying theory infants should maintain at least an approximate sense of the quantity of hidden objects. Understanding the robustness of infants' representational failures hence provides an important test-case for potentially adjudicating between these views.

Beyond representations of sets of objects, prominent theories of children's number word acquisition also rely on the "two systems" theory of numerical representation. These theories suggest that the stage-like developmental progression in children's number word understanding reflects the catastrophic set size limitation on infants' object representations (Carey, 2009; Carey & Barner, 2019; Le Corre & Carey, 2007; vanMarle, 2015). According to this view, children's capacity to track one, two, or three objects in parallel allows them to form set-based representations of "oneness," "twoness," and "threeness." Typically, over an extended period between 2 and 4 years of age, children gradually learn to map the corresponding number words (i.e., "one," "two," "three") to these set-based representations in a stage-like fashion—first understanding "one" but nothing more, and then gradually acquiring "two" over the next few months, and then gradually acquiring "three" in a similar way (Wynn, 1990, 1992). However, this piecemeal acquisition process stops sometime after children acquire "three." Instead of continuing to map each larger numerosity to their corresponding number word, children seem to make a "conceptual leap" and induce that the counting sequence can be used to generate precise and large numerical quantities. This qualitative shift in the number word acquisition process is attributed to the limited capacity children have to keep track of four or more objects in parallel—children may be running out of the capacity to form set-based representations of four or more objects, which motivates them to adopt one-to-one correspondence and the counting sequence to help keep track of larger quantities in a precise way (Carey, 2009; Carey & Barner, 2019). In contrast to this "object-based" theory of number word acquisition, others have proposed that children form noisy representations of numerosities throughout the number word acquisition process, without ever being constrained by an object tracking capacity limit (e.g., Gelman & Gallistel, 1978; Gallistel & Gelman, 2000; Wagner & Johnson, 2011). Quantifying the strength of the evidence for infants' representational capacity limits can provide important clarity to this debate.

Theories of the development of object representational capacities in working memory also must account for infants' catastrophic set size limits, since such catastrophic failures do not appear to be a feature of object working memory for older children or adults. If catastrophic set size limit effects are robust in infancy, it would suggest that the way infants represent objects in working memory is qualitatively different from the way older children and adults do, and would therefore reflect a structural limitation over working memory in infancy that is subsequently overcome with development or with the acquisition of more efficient working memory strategies. However, if catastrophic set size limit effects are not robust, the existence of these failures in infant experiments could suggest that infants' working memory may store more graded representations of objects (Munakata, 2001) or may have a finite amount of visual working memory resource to devote to representations of objects (Bays & Husain, 2008), and when these resources are spread too thin, the result is a set of unreliable representations of hidden objects, rather than the complete obliteration of these representations. That is, infants' apparent failures when more than four objects are hidden could reflect representational "fading" or representational noise rather than catastrophic representational failure. If this is the case, we should observe insufficient evidence for catastrophic failures, because infants' representational noise should lead to inconsistent performance across studies, instead of consistent failures across studies.

2.2 | How strong is the evidence for the "catastrophic" set size limit?

While many studies have demonstrated catastrophic set size limits in infants, it remains unclear just how reliable catastrophic failure effects are. In general, null results are difficult to interpret on an individual-study level, for several reasons. First, the conclusions that are drawn from these studies are based on the results of null hypothesis significance testing, which does not yield the probability of the null hypothesis, but instead gives the probability of the data if it were the case that the null hypothesis is true (Gallistel, 2009; Gigerenzer et al., 2004; Tyron, 1998). This means that the extent to which the results provide evidence *in favor of* the null hypothesis—that infants fail to represent even an approximate the number of objects in the hidden array—is not known.

Further, the problem of interpreting null results in infant studies that show catastrophic set size limits is exacerbated by the fact that, in the typical infant experiment described in the literature above, both the number of subjects and the number of trials per subject are extremely low. Obtaining robust evidence for a null effect requires more participants and/or more trials than obtaining evidence for a non-null effect (Lakens, 2017). Studies that used the foraging method typically had samples of $n = 16$ infants per experiment and only one two-alternative-forced choice trial per infant. In practice, this means that a "success" or "failure" could hinge on the behavior of a single infant. For example, if 13 out of 16 infants in a foraging task chose the container with more crackers, this would be interpreted as a "success" (binomial test $p = 0.02$). However, if 12 out of 16 infants chose the container with more crackers, this would be interpreted as a "failure" ($p = 0.08$), and in the literature described above failures and successes were rarely statistically compared. Studies that use the manual search method have more trials (usually two trials in which infants are expected to "fail" contrasted with two trials in which infants are expected to "succeed") but still typically include around $n = 16$–18 infants per experiment. This makes it difficult to assess the strength of these null results on an experiment-by-experiment basis.

Given the significance of catastrophic set size limits for theories of development, understanding how reliable and robust these effects are is critical. Further, on a more practical level, if it is the case that infants' representations of objects are mostly wiped out in the presence of four or more objects, this poses a significant information processing challenge for infants, who need to maintain a continuous representation of the three-dimensional world as they navigate their environments and learn from others. This information processing limitation should be taken into account when posing theories of infants' visual processing and learning across domains. Hence, quantifying the robustness of the evidence for infants' catastrophic set size limits can provide a key data point for theories of the cognitive mechanism underlying how infants represent and learn about the world.

3 | BAYESIAN META-ANALYSIS

To examine the robustness of the evidence for catastrophic set size limits in infants, we conducted a Bayesian meta-analysis. Using a Bayesian approach to meta-analysis is ideal for infant data: it is robust to low n and can quantify the strength of the null results by giving the odds of the null hypothesis over the alternative given the extant data (Bartoš et al., 2023; Berkhout et al., 2023). The goal is to use all of the evidence from the literature to estimate the extent to which the evidence points toward the null hypothesis that infants' object representational capacities fail "catastrophically" when they are tasked with tracking four or more objects in a location.

3.1 | Method

We followed the "Preferred Reporting Items for Systematic Reviews and Meta-analysis" (PRISMA) statement (Moher et al., 2009) when preparing the meta-analysis. The preregistration and all data associated with this meta-analysis can be found at https://osf.io/9cpsx/.

3.1.1 | Report identification

Based on author knowledge, we identified 13 papers published to date that measured infants' ability to represent multiple, identical hidden objects. We obtained one unpublished manuscript through personal communication with the authors. We posted a request for unpublished data on the listservs for the Cognitive Development Society and the International Congress of Infant Studies which yielded no relevant results. We then conducted a Google Scholar search using the Publish or Perish tool (Harzing, 2007) on May 18, 2023 for articles that were not citations or patents. We used the keywords *infant "memory for hidden objects"* (60 reports) and *infant "object tracking"* (995 reports). We also conducted a search directly on Google Scholar for articles that cited Feigenson et al. (2002) with the keyword *infant* (764 reports). This yielded a total of 1819 reports, including duplicates across search terms.

3.1.2 | Report selection

We selected reports based on the following criteria:

1. Participants were typically developing infants under 22 months of age, since previous work suggested that infants beyond 22 months may begin to use the singular-plural markers of formal language to overcome catastrophic set size limits (e.g., Barner et al., 2007).
2. Participants were tested using either a manual search task or a foraging task[1].
3. Participants observed at least four and not more than six identical objects being hidden, since larger quantities are likely represented using an analog-magnitude-like representational format, rather than using individual object representations (vanMarle & Wynn, 2011; Xu, 2003).[2]
4. No cues that could facilitate infants' ability to concurrently represent multiple objects were present (e.g., chunking, counting).

We thus excluded reports that used different paradigms, such as looking time studies (e.g., Kibbe & Feigenson, 2016; Wang, 2023), as well as experiments from reports that included chunking or counting cues (e.g., Experiments 1–4 in Feigenson & Halberda, 2004).

This left us with 12 reports in total, with a total of 22 relevant experiments that met the inclusion criterion (Table 1), drawn from three different labs in the United States.

3.1.3 | Data entry

We created a database for the meta-analysis to include all 22 experiments in the 12 reports. For each experiment, we recorded an effect size. Effect sizes from the manual search tasks were calculated in *Cohen's d* based on the reported t statistics:

$$d = \frac{t}{\sqrt{n}}$$

And effect sizes from the foraging tasks were converted to *Cohen's d* by first calculating the Log Odds Ratio (LogOR):

$$LogOR = Log\left(\frac{p}{1-p}\right)$$

Then, converting the LogOR to Cohen's d:

$$d = LogOR \times \frac{\sqrt{3}}{\pi}$$

Each experiment was also coded for the authors, year published, mean age and age range (in months), task (manual search or foraging), and sample size. Confidence intervals of effect size *d* were calculated using the d.ci function in the psych R package (Revelle, 2023).

3.2 | Analysis and results

The goal of our analysis was to test the robustness of the evidence for infants' catastrophic representational failures by measuring the strength of the null effect that infants cannot concurrently represent four or more individual hidden objects. We conducted a Robust Bayesian Meta–Analysis using the RoBMA package (Bartoš & Maier, 2020) in R (R Core Team, 2022). The Robust Bayesian Meta–Analysis simultaneously constructs 36 different models that assume the presence or absence of effect, as well as different ways to adjust for the heterogeneity of effect sizes and the presence of publication bias, and computes an Inclusion Bayes Factor for the presence of effect (i.e., the evidence in favor of the alternative hypothesis) using Bayesian Model Averaging (see Bartoš et al., 2023 for more details). Critical for our current meta-analysis, the calculated inclusion Bayes Factor distinguishes between "absence of evidence" (i.e., Inclusion BF is close to 1, lacking evidence for either null or alternative hypothesis) from "evidence of absence" (i.e., Inclusion BF is close to 0, supporting the null hypothesis). Although Bayes factors are a continuous measure and contain more information than traditional null hypothesis significance testing or simple categorizations of statistical results, to facilitate interpretation, we will report both the Bayes factors and categorical

TABLE 1 Reports included in the current meta-analysis for infants' ability to concurrently represent four to six objects.

Authors	Year	Ages (months)	Task	# Experiments	Total # participants	Peer reviewed
Barner, Thalwitz, Wood, Yang, & Carey	2007	18; 20	Manual Search	2	33	Yes
Feigenson & Carey	2003	12–14	Manual Search	1	16	Yes
Feigenson & Carey	2005	10–12	Foraging; Manual Search	2	32	Yes
Feigenson & Halberda	2004	14	Manual Search	1	16	Yes
Feigenson, Carey, & Hauser	2002	10; 12	Foraging	4	64	Yes
Rosenberg & Feigenson	2013	14	Manual Search	1	22	Yes
Rosenberg & Feigenson	in prep	14	Manual Search	1	24	No
Stahl & Feigenson	2014	16	Manual Search	2	32	Yes
Stahl & Feigenson	2018	16	Manual Search	2	32	Yes
Stahl, Pareja, & Feigenson	2023	16	Manual Search	1	16	Yes
vanMarle	2013	10–12	Foraging	1	16	Yes
Wang & Feigenson	2019	14–20	Manual Search	4	64	Yes
Total				***22***	***367***	

interpretations (Jeffreys, 1939), regarding Bayes factors between 1 and 3 as weak or anecdotal evidence for the alternative hypothesis (or between 0.33 and 1 for the null hypothesis), between 3 and 10 (or between 0.1 and 0.33) as moderate evidence, and larger than 10 (or smaller than 0.1) as strong evidence.

First, following our preregistered analysis plan, we examined our full dataset, including studies using manual search and foraging tasks. We used the effect size for infants' performance when tracking four to six hidden objects and calculated confidence intervals as input for the RoBMA function (Bartoš & Maier, 2020) in R (R Core Team, 2022). The model averaged effect size estimate was −0.001 (95% CI [−0.020, 0.070]; see Figure 2 for forest plot). The posterior model probability was 0.090, *Inclusion BF* = 0.099. We also conducted a sequential analysis (not preregistered) using the Bayesian Meta Analysis engine in JASP (Love et al., 2019), which revealed that the posterior probabilities of the null hypotheses outperformed the alternative hypotheses around the inclusion of the fifth study (Figure 3). These results offer strong support for the evidence that infants fail to concurrently represent four to six hidden objects.

Second, following our preregistered plan, we separately analyzed the set of studies that used the manual search task and the set of studies that used the foraging task. There were 16 effect sizes from 10 reports using the manual search task, yielding a model averaged effect size estimate *Cohen's d* = 0 (95% CI [−0.075, 0.128]) with a posterior model probability of 0.139 and *Inclusion BF* = 0.162, indicating moderate support for the null hypothesis that infants fail to concurrently represent sets of four to six hidden objects. There were six effect sizes from three reports using the foraging task, yielding a model averaged effect size estimate of −0.014 (95% CI [−0.202, 0.018]), with a posterior model probability of 0.125 and *Inclusion BF* = 0.143, indicating moderate support for the null hypothesis.

Last, we conducted a robust Bayesian meta-analysis on effect sizes from experiments measuring infants' ability to concurrently represent sets *within the set size limit*—that is, sets of *fewer than four* objects, using either manual search or foraging tasks (this analysis was not preregistered). We identified 30 relevant experiments (24 manual search and 6 foraging) from 12 reports that we obtained using our preregistered search criteria. These reports are detailed in Table S1. The RoBMA model-averaged effect size was 0.497 (95% CI [0.000, 0.712], Figure S1), *posterior probability* = 0.875, *Inclusion BF* = 6.973 (see Figure S2 for sequential analysis). These results offer moderate support for infants' success to concurrently represent fewer than four hidden objects.

4 | GENERAL DISCUSSION

We aimed to quantify the evidence in favor of the null hypothesis that infants fail to represent objects when tasked with representing a set of more than three objects. We conducted a Robust Bayesian Meta-Analysis on experiments that reported observing infants' "catastrophic" set size limits in tasks that required them to keep track of four to six hidden physical objects. The results of our Bayesian meta-analysis overall supported the null hypothesis, yielding odds of roughly 10 to 1 in favor of the null over the alternative hypothesis. That is, the evidence strongly supports the notion that infants' fail to concurrently represent sets of four to six occluded objects. Our preregistered exploratory analyses, which examined the manual search and foraging tasks separately, showed that manual search tasks appeared to yield stronger evidence for the null than foraging tasks. This result is perhaps not surprising given that there were fewer studies using the foraging task than the manual search task in our analysis, and given that the foraging task involves only a single trial binary measure compared to the

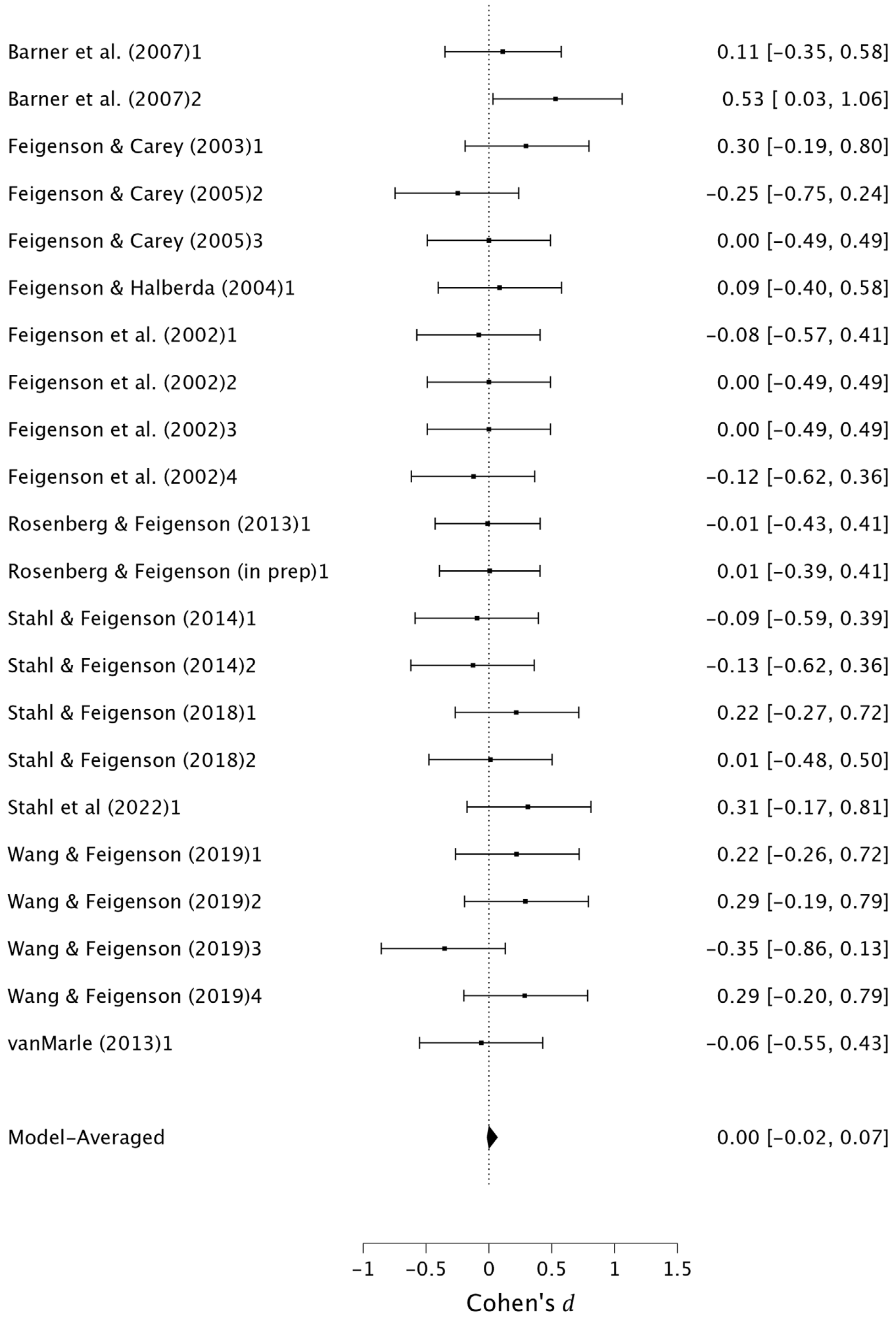

FIGURE 2 Forest plot with point estimates of effect sizes from studies that measured infants' ability to concurrently represent sets of four to six objects. Bars represent 95% CI.

multiple-trial continuous measure obtained in manual search. Nevertheless, the totality of the evidence supports the null hypothesis. We also conducted a complementary analysis on results from experiments examining infants' ability to concurrently represent sets of *fewer than four* objects (i.e., within the set size limit), and found that the evidence supported their ability to do so. Below, we discuss the implications of these results and suggest avenues for future research.

4.1 | What can these results tell us about the format of infants' representations of objects?

Feigenson et al. (2002) argued that infants' catastrophic set size limits can be explained by early limitations in the representational capacity of the object-file system (see Leslie et al., 1998). At the time Feigenson and colleagues were conducting their studies on infants' object

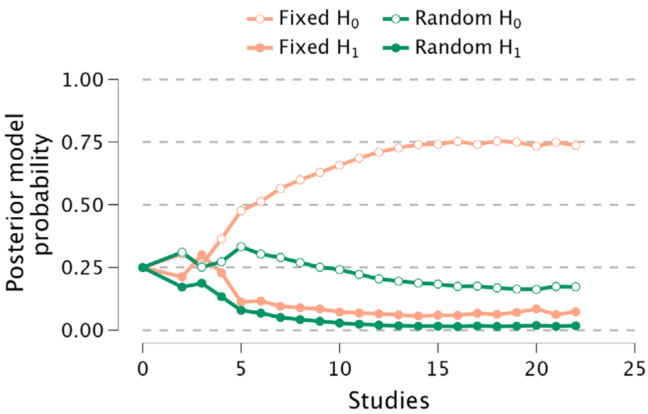

FIGURE 3 Sequential analysis of the posterior model probabilities for the Bayesian meta-analysis on infants' ability to concurrently represent sets of four to six objects.

representational capacities, there was a flurry of research examining limitations on adults' ability to visually track multiple objects simultaneously (e.g., Scholl & Pylyshyn, 1999). This work seemed to show some parallels with infants' pattern of behaviors in manual search and foraging tasks. For example, adults who were tasked with simultaneously tracking the trajectories of multiple moving dots in a computer display were able to do so only when the number of dots they were tasked with tracking was four or fewer. If they were tasked with tracking greater than four objects, adults had difficulty tracking even a subset of the objects, typically tracking only one or none of the moving set (Scholl & Pylyshyn, 1998).

Feigenson and Carey (2005) hypothesized that infants, like adults, may deploy discrete object files to track the trajectory of individual objects through space even into occlusion. However, unlike adults, they may be unable to strategically attend to only a subset of the objects when their object representational capacities are overwhelmed. Indeed, when sets are structured to facilitate perceptual or conceptual grouping of objects into subsets, infants are better able to represent sets of four objects, and no longer display the signatures of catastrophic representational failures (e.g., Feigenson & Halberda, 2008; Kibbe & Feigenson, 2016; Rosenberg & Feigenson, 2013; Stahl & Feigenson, 2014, 2018; Stahl et al., 2023; see also Wang & Feigenson, 2019; for review, see Kibbe & Stahl, 2023). However, others have suggested that infants' object representations may persist in a graded or faded fashion, and failures thus are indicative of representational fuzziness due to task demands rather than a complete loss of a discrete object representation (e.g., Munakata, 2001). The results of our Bayesian meta-analysis yielded evidence that, when infants are tasked with tracking sets of objects into occlusion, infants fail to maintain even a subset of the hidden objects or the approximate amount of the total set. The evidence in favor of the null supports the idea that infants' representations of objects consist of discrete, all-or-nothing representations rather than graded or faded representations of individual objects.

However, several open questions remain. First, when infants' object representations fail catastrophically, what (if anything) remains? Although it is possible that infants are left with "absolutely nothing"—as if no objects were hidden ever—we suspect that this is unlikely. While the experimental evidence is limited, one small study (Feigenson and Carey (2005) Experiment 2) tested infants in a foraging task in which four objects were hidden in one container and either one object (1 vs. 4 condition, $n = 16$) or zero objects (0 vs. 4 condition, $n = 16$) were hidden in the other container (in the 0 vs. 4 condition, the experimenter waved her hand over the empty location to equate attention and movement across the two locations). They replicated infants' failure in the 1 versus 4 condition (8/8 infants crawled to the location containing 4 crackers), but found that infants crawled to the location containing four crackers in the 0 versus 4 condition at rates above chance (13/16 infants crawled to the location containing 4 crackers). However, infants' crawling behavior between these two conditions was not significantly different (Fisher's exact test $p = 0.14$, an analysis that was not included in Feigenson & Carey, 2005). The results of this study may suggest that infants are representing that "something" was in the location in which four crackers were hidden, with the important caveat that this single experimental result should be interpreted with a considerable amount of caution. Nevertheless, we suggest that future studies using similar comparisons could potentially replicate and extend this result with larger samples, and we encourage additional work that aims to understand the representational format of "forgotten" sets of four or more objects.

Second, *the point at which* infants' representations fail catastrophically is not known. One possibility is that catastrophic failure occurs at the point at which the objects are occluded; that is, once infants lose visual access to the arrays, they are unable to store representations of the objects in working memory because *working memory itself is limited* (see Kibbe, 2015; Zosh & Feigenson, 2009). Another possibility is that infants fail to represent sets of more than three of fully visible three-dimensional objects prior to memory storage. Indeed, adults in multiple object tracking tasks often are unable to track fully visible objects when tasked with tracking more than three or four objects, suggesting that tracking objects—even when those objects are not occluded—comes with its own limit. Still another possibility is that failures occur when infants attempt to *act on* their representations—that is, at the point of "retrieval" of the representations from working memory to make decisions about searching, choosing, or looking. Future work should attempt to tease apart these possibilities.

4.2 | What can these results tell us about the format of infants' representations of quantities?

According to the "unifying theory," behavioral differences in people's responses to small versus large numerical quantities are only reflective of quantitative differences in the amount of internal noise when representing quantities—which exist for both small and large sets (Cheyette & Piatadosi, 2020). This theory receives support from studies in which infants track arrays of two-dimensional objects from one static frame to another, such as those using the Change Detection Paradigm. Specifically, infants can successfully keep track of both small and large quantities of objects, and their performance in both obeys Weber's Law (Starr et al., 2013). Results like this are consistent with the "unifying theory" of numerical representation, which suggests that there is no fundamental divide between small versus large quantities in terms of representational format.

In contrast, the "two systems theory" suggests a qualitative difference between representations of small versus large sets of objects (Feigenson et al., 2004). Support for this theory comes from studies that found distinct behavioral patterns in infants when representing small versus large arrays of 2D objects (e.g., Xu, 2003). Similarly, the infant brain shows distinct neural signatures in response to small versus large quantities of objects (Hyde & Spelke, 2011). Results from the current Bayesian meta-analysis provide additional support for the "two systems" account by quantifying the robustness of the evidence that infants fail to represent any approximate trace of four or more objects.

What accounts for the apparent differences across studies? It has been theorized that differences in attentional demands in different tasks can explain these divergent findings (Hyde, 2011). Specifically, when items are presented in a context where it is relatively easy to track each individual, they will be represented with object files and exhibit a set-size limit (see also Kibbe, 2015); conversely, when there is too much attentional demand such that it is hard to track each individual, for example, when too many items flash too quickly, the items will be represented as an ensemble instead of a distinct set of individuals.

The current meta analysis results are consistent with half of this theory that items presented in the manual search and foraging tasks are represented as individuals and thus are subject to the "catastrophic" set-size limit. However, according to a recent meta-analysis on infants' representation of large numerical arrays, existing results on infants' success when tracking large numerical arrays do not provide decisive evidence that infants are indeed tracking large approximate quantities successfully (Smyth & Ansari, 2020), leaving it unclear whether a qualitative shift exists for infants' tracking of small versus large arrays. Taken together, these results suggest that the field needs large-scale replications of the findings on infants' object tracking to further adjudicate between theories of the representational format of objects.

4.3 | What can these results tell us about children's acquisition of number words?

Although the current meta-analysis only included studies of precounting infants, it provides strong support for the existence of an object tracking system that is subject to a catastrophic set size limit, which sets the stage for theories of learning that rely on such a capacity-limited cognitive system. For example, an ongoing debate about children's number word acquisition is whether this object tracking system serves as the critical foundation for number word learning, without any support from children's ability to represent approximate quantities (Carey, 2009; Carey & Barner, 2019; Gallistel & Gelman, 2000; Wagner & Johnson, 2011). However, our results only provide support for the existence of a set size limit at about four in our ability to track individual objects. The theory that this set size limit corresponds to the stage-like developmental trajectory of number word acquisition, which also has an apparent qualitative shift at about "four," requires further evidence.

On the one hand, it is plausible that children's representation of "one-ness" and "two-ness" relies on the same object tracking system as infants', which reaches its capacity limit at "four-ness" or above, and this limit therefore pushes children toward an induction of the Cardinal Principle (Carey, 2009). On the other hand, unlike infants, who seem to lose almost every trace of the object array when the object tracking system reaches its set size limit (Feigenson & Carey, 2005), older children and adults can keep track of a subset of the hidden objects (Blankenship et al., 2020; Scholl & Pylyshyn, 1999; Trick et al., 2005). Moreover, adults can also represent the approximate cardinality of the set in a multiple object tracking task, even when the set size exceeds their tracking capacity (Ma & Flombaum, 2013). It is likely that once the object tracking system reaches its capacity limit for sets of four or more objects, older children still have access to the approximate cardinality of the set. And this shift in representational format—from having precise tracking of each object in a set, to having an approximate representation of the overall cardinality of the set—may contribute to children's induction of the Cardinal Principle. Under this account, approximate number representations may also play a role in children's acquisition of number words.

Just like we know little about the cognitive mechanisms underlying infants' catastrophic failure when concurrently representing four or more objects, we know little about what enables older children to not experience catastrophic failure, or whether older children have access to approximate representations of large sets of objects. To speculate, if older children can "fall back" on approximate numerical representations of larger object sets when their object tracking system fails, they should be able to rely on these numerical representations when interpreting large number words such as "five" or "ten." However, existing research has yielded mixed results regarding children's representation of quantities beyond their number word knowledge. While some studies found that children fail to map larger number words to larger quantities (Le Corre & Carey, 2007; Slusser & Sarnecka, 2011), others found that children can indeed map larger number words to larger quantities prior to acquiring the Cardinal Principle (Gunderson et al., 2015; Odic et al., 2015; Wagner et al., 2019). It is possible that children have to gradually acquire the ability to overcome the catastrophic set size limit and access the approximate cardinality of the set. Wang and Feigenson (2019) suggest that counting may be one of the mechanisms that can help infants overcome the catastrophic set size limit of their object representations, and access their approximate numerical representation of the set. Additionally, Shusterman et al. (2016) found that acquisition of the Cardinal Principle coincides with an improvement in children's numerical approximation precision. Further work is needed to examine what role the object tracking system plays in children's learning of number words, and what role number word acquisition plays in shaping how children represent sets of objects.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and preregistration document are available at https://osf.io/9cpsx/

ORCID

Jinjing (Jenny) Wang https://orcid.org/0000-0001-8350-7656
Melissa M. Kibbe https://orcid.org/0000-0002-9088-2523

ENDNOTES

[1] We opted not to include violation-of-expectation (VOE) looking time studies (two experiments from two reports) in our preregistered confirmatory analysis. We reasoned that the effect sizes may not be comparable since VOE requires only passive viewing rather than active behavior. Including the two additional experiments did not qualitatively change the results, *posterior model probability* = 0.198, *Inclusion BF* = 0.246.

[2] We identified two additional experiments from one report where infants had to track eight objects in a foraging task (vanMarle, 2013). Including these two additional experiments did not qualitatively change the results, *posterior model probability* = 0.098, *Inclusion BF* = 0.109.

REFERENCES

Applin, J. B., & Kibbe, M. M. (2019). Six-month-old infants predict agents' goal-directed actions on occluded objects. *Infancy*, *24*(3), 392–410. https://doi.org/10.1111/infa.12282

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

Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. *Developmental Psychology*, *23*(5), 655. https://doi.org/10.1037/0012-1649.23.5.655

Barner, D., Thalwitz, D., Wood, J., Yang, S. J., & Carey, S. (2007). On the relation between the acquisition of singular–plural morpho-syntax and the conceptual distinction between one and more than one. *Developmental Science*, *10*(3), 365–373. https://doi.org/10.1111/j.1467-7687.2007.00591.x

Bartoš, F., & Maier, M. (2020). RoBMA: An R package for robust Bayesian meta-analyses. R package version 3.0.1. https://CRAN.R-project.org/package=RoBMA

Bartoš, F., Maier, M., Wagenmakers, E. J., Doucouliagos, H., & Stanley, T. D. (2023). Robust Bayesian meta-analysis: Model-averaging across complementary publication bias adjustment methods. *Research Synthesis Methods*, *14*(1), 99–116. https://doi.org/10.1002/jrsm.1594

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Berkhout, S. W., Haaf, J. M., Gronau, Q. F., Heck, D. W., & Wagenmakers, E. J. (2023). A tutorial on Bayesian model-averaged meta-analysis in JASP. *Behavior Research Methods*, 1–23.

Blankenship, T. L., Strong, R. W., & Kibbe, M. M. (2020). Development of multiple object tracking via multifocal attention. *Developmental Psychology*, *56*(9), 1684. https://doi.org/10.1037/dev0001064

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology*, *44*(4), 388–426.

Bower, T. G. R. (1967). The development of object-permanence: Some studies of existence constancy. *Perception & Psychophysics*, *2*, 411–418.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835. https://doi.org/10.1016/j.tics.2019.07.004

Cordes, S., & Brannon, E. M. (2008). Quantitative competencies in infancy. *Developmental Science*, *11*(6), 803–808.

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*(2), 239–264.

Cheng, C., & Kibbe, M. M. (2022). Development of updating in working memory in 4–7-year-old children. *Developmental Psychology*, *58*(5), 902–912. https://doi.org/10.1037/dev0001337

Cheng, C., & Kibbe, M. M. (2024). Children's use of reasoning by exclusion to infer objects' identities in working memory. *Journal of Experimental Child Psychology*, *237*, 105765. https://doi.org/10.1016/j.jecp.2023.105765

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, *4*(12), 1265–1272. https://doi.org/10.1038/s41562-020-00946-0

Feigenson, L. (2008). Parallel non-verbal enumeration is constrained by a set-based limit. *Cognition*, *107*(1), 1–18. https://doi.org/10.1016/j.cognition.2007.07.006

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, *97*(3), 295–313. https://doi.org/10.1016/j.cognition.2004.09.010

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*(2), 150–156. https://doi.org/10.1111/1467-9280.00427

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Feigenson, L., & Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition*, *91*(2), 173–190.

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, *116*(2), 439. https://doi.org/10.1037/a0015251

Gallistel, C. R. (2020). Getting numbers into brains. *Nature Human Behaviour*, *4*(12), 1222–1223. https://doi.org/10.1038/s41562-020-00970-0

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, *4*(2), 59–65. https://doi.org/10.1016/S1364-6613(99)01424-2

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Harvard University Press.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual. *The Sage Handbook of Quantitative Methodology for the Social Sciences*, 391–408.

Gross, H. J., Pahl, M., Si, A., Zhu, H., Tautz, J., & Zhang, S. (2009). Number-based visual generalisation in the honeybee. *PloS ONE*, *4*(1), e4263. https://doi.org/10.1371/journal.pone.0004263

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35–55.

Harzing, A. W. (2007). Publish or perish. Harzing. com. Research in International Management software release, accessed August, 27.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi–free–ranging rhesus monkeys. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *267*(1445), 829–833. https://doi.org/10.1098/rspb.2000.1078

Hauser, M. D., & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*(4), 367–401. https://doi.org/10.1016/S0010-0285(03)00050-1

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience*, *5*, 150. https://doi.org/10.3389/fnhum.2011.00150

Jeffreys, H. (1939). *The theory of probability*. OuP Oxford.

Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what' and 'where' information. *Developmental Science*, *6*(3), 360–373. https://doi.org/10.1111/1467-7687.00290

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*(2), 153–177. https://doi.org/10.1016/j.cognition.2004.09.009

Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science*, *24*(6), 433–439. https://doi.org/10.1177/0963721415605831

Kibbe, M. M., & Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition*, *146*, 251–263. https://doi.org/10.1016/j.cognition.2015.09.022

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. https://doi.org/10.1177/0956797611420165

Kibbe, M. M., & Leslie, A. M. (2016). The ring that does not bind: Topological class in infants' working memory for objects. *Cognitive Development*, *38*, 1–9. https://doi.org/10.1016/j.cogdev.2015.12.001

Kibbe, M. M., & Leslie, A. M. (2019). Conceptually rich, perceptually sparse: Object representations in 6-month-old infants' working memory. *Psychological Science*, *30*(3), 362–375. https://doi.org/10.1177/0956797618817754

Kibbe, M. M., & Stahl, A. E. (2023). Objects in a social world: Infants' object representational capacity limits are shaped by objects' social relevance. *Advances in Child Development & Behavior*, 69–97.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438.

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.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., ŠmÍra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Ma, Z., & Flombaum, J. I. (2013). Off to a bad start: Uncertainty about the number of targets at the onset of multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1421.

Mack, W. (2006). Numerosity discrimination: Infants discriminate small from large numerosities. *European Journal of Developmental Psychology*, *3*(1), 31–47. https://doi.org/10.1080/17405620500347695

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*, T.. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Mou, Y. (2014). Two core systems of numerical representation in infants. *Developmental Review*, *34*(1), 1–25. https://doi.org/10.1016/j.dr.2013.11.001

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7), 309–315. https://doi.org/10.1016/S1364-6613(00)01682-X

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241. https://doi.org/10.1037/1082-989X.5.2.241

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138*, 102–121.

Piaget, J. (1954). The development of object concept (M. Cook, Trans.). In J. Piaget & M. Cook (Trans.), *The construction of reality in the child* (pp. 3–96). Basic Books. https://doi.org/10.1037/11168-001

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, *50*(1-3), 363–384. https://doi.org/10.1016/0010-0277(94)90036-1

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Revelle, W. (2023). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.4.1, https://CRAN.R-project.org/package=psych

Rosenberg, R. D., & Feigenson, L. (2013). Infants hierarchically organize memory representations. *Developmental Science*, *16*(4), 610–621.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259–290. https://doi.org/10.1006/cogp.1998.0698

Shinskey, J. L., & Munakata, Y. (2005). Familiarity breeds searching: Infants reverse their novelty preferences when reaching for hidden objects. *Psychological Science*, *16*(8), 596–600. https://doi.org/10.1111/j.1467-9280.2005.01581.x

Smyth, R. E., & Ansari, D. (2020). Do infants have a sense of numerosity? A p-curve analysis of infant numerosity discrimination studies. *Developmental science*, *23*(2), e12897. https://doi.org/10.1111/desc.12897

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605. https://doi.org/10.1037/0033-295X.99.4.605

Stahl, A. E., & Feigenson, L. (2014). Social knowledge facilitates chunking in infancy. *Child Development*, *85*(4), 1477–1490. https://doi.org/10.1111/cdev.12217

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Stahl, A. E., & Feigenson, L. (2018). Infants use linguistic group distinctions to chunk items in memory. *Journal of Experimental Child Psychology*, *172*, 149–167. https://doi.org/10.1016/j.jecp.2018.03.005

Stahl, A. E., Pareja, D., & Feigenson, L. (2023). Early understanding of ownership helps infants efficiently organize objects in memory. *Cognitive Development*, *65*, 101274. https://doi.org/10.1016/j.cogdev.2022.101274

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy*, *18*(6), 927–941. https://doi.org/10.1111/infa.12008

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "Catch the Spies" task. *Cognitive Development*, *20*(3), 373–387. https://doi.org/10.1016/j.cogdev.2005.05.009

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological review*, *101*(1), 80. https://doi.org/10.1037/0033-295X.101.1.80

Tyron, W. W. (1998). The inscrutable null hypothesis. *American Psychologist*, *53*(7), 796. https://doi.org/10.1037/0003-066X.53.7.796.b

Uller, C., & Lewis, J. (2009). Horses (*Equus caballus*) select the greater of two quantities in small numerical contrasts. *Animal Cognition*, *12*, 733–738. https://doi.org/10.1007/s10071-009-0225-0

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (*Plethodon cinereus*) go for more: Rudiments of number in an amphibian. *Animal Cognition*, *6*, 105–112. https://doi.org/10.1007/s10071-003-0167-x

Van de Walle, G. A., Carey, S., & Prevor, M. (2000). Bases for object individuation in infancy: Evidence from manual search. *Journal of Cognition and Development*, *1*(3), 249–280. https://doi.org/10.1207/S15327647JCD0103_1 vanMarle, K. (2013). Infants use different mechanisms to make small and large number ordinal judgments. *Journal of Experimental Child Psychology*, *114*(1), 102–110. https://doi.org/10.1016/j.jecp.2012.04.007

vanMarle, K. (2015). Foundations of the formal number concept: How preverbal mechanisms contribute to the development of cardinal knowledge. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Evolutionary origins and early development of number processing* (pp. 175–199). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-420133-0.00007-7

vanMarle, K., & Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Developmental Science*, *14*(3), 502–515. https://doi.org/10.1111/j.1467-7687.2010.00998.x

Wagner, K., Chu, J., & Barner, D. (2019). Do children's number words begin noisy? *Developmental Science*, *22*(1), e12752. https://doi.org/10.1111/desc.12752

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. https://doi.org/10.1016/j.cognition.2010.11.014

Wang, J. J. (2023). Does virtual counting count for babies? Evidence from an online looking time study. *Developmental Psychology*, *59*(4), 669. https://doi.org/10.1037/dev0001478

Wang, J., & Feigenson, L. (2019). Infants recognize counting as numerically relevant. *Developmental Science*, *22*(6), e12805. https://doi.org/10.1111/desc.12805

Wang, J., & Feigenson, L. (2023). What aspects of counting help infants attend to numerosity? *Infancy*, *28*(2), 218–239. https://doi.org/10.1111/infa.12512

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*(2), 125–166. https://doi.org/10.1016/S0010-0277(99)00035-9

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193.

Wynn, K. (1992). Addition and subtraction in human infants. *Nature*, .

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. https://doi.org/10.1016/S0010-0277(03)00050-7

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153.

Zosh, J. M., & Feigenson, L. (2009). Beyond 'what' and 'how many': capacity, complexity and resolution of infants' object representations. *The Origins of Object Knowledge*, 25–51. https://doi.org/10.1093/acprof:oso/9780199216895.003.0002

Zosh, J. M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *Journal of Experimental Child Psychology*, *113*(3), 322–336. https://doi.org/10.1016/j.jecp.2012.07.005

Zosh, J. M., & Feigenson, L. (2015). Array heterogeneity prevents catastrophic forgetting in infants. *Cognition*, *136*, 365–380. https://doi.org/10.1016/j.cognition.2014.11.042

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, J. J., & Kibbe, M. M. (2024). "Catastrophic" set size limits on infants' capacity to represent objects: A systematic review and Bayesian meta-analysis. *Developmental Science*, e13488. https://doi.org/10.1111/desc.13488