COMMENTARIES

Evidence for numerical abilities in young infants: a fatal flaw?

Susan Carey

Harvard University, USA

Wynn's (1992) well-known and oft-replicated[1] *Nature* paper established that 4-month-old infants look longer at impossible than at possible outcomes in 'addition' and 'subtraction' events. In a 1 + 1 addition event, for instance, the infant was shown a single object on a stage, which was then covered by a screen. A second object was then introduced behind the screen and the screen then was lowered, revealing either the possible outcome of 2 objects or an impossible outcome of 1 object. In this 1 + 1 = 2 or 1 event, infants looked longer at the impossible outcome of 1 object. Similarly, infants look longer at the impossible outcome in a 2 – 1 = 2 or 1 subtraction event or a 1 + 1 = 2 or 3 addition event.

Wynn took these findings to show that very young infants symbolically represent the number of hidden objects behind screens, and can operate on symbolic representations of number by adding or subtracting 1, yielding number representations of the resulting sets. She suggested that the format of representation may be the analog magnitudes first suggested by Meck and Church (1983) (see Dehaene, 1997, for evidence that non-human animals and adult humans deploy analog magnitude representations of number).

Cohen and Marks' paper has three main points. First, and most important, they pose a strong empirical challenge to Wynn's interpretation of the data from the addition/subtraction studies. Second, they argue that familiarity preferences may contribute to patterns of looking times in such studies. Third, they call for caution when attributing sophisticated abilities, such as the capacity to add and subtract, to very young infants.

Although Cohen and Marks' remarks are aimed at Wynn's proposals, their data bear equally on the other major interpretation of the infant addition/subtraction data, called the 'object file' model by Uller, Huntley-Fenner, Carey and Klatt (1999). According to the object file model, infants establish a representation of each individual object behind the screen, updating these representations as objects are added to or subtracted from the set (see also Scholl & Leslie, 1999; Simon, 1997). The representation of the updated hidden set is compared to the set of objects revealed when the screen is removed on the basis of 1–1 correspondence, a computation that established numerical equivalence (but see Feigenson, Carey & Spelke, in press, and Feigenson, Carey & Hauser, in press).

Cohen and Marks' (this volume) lovely paper provides a strong challenge to both the analog magnitude and the object file interpretations of the infant addition/subtraction data. Their data challenge whether infants are updating representations of hidden sets in a manner that involves any kind of addition and subtraction. Rather, Cohen and Marks suggest that the infant addition/subtraction data may be accounted for in terms of a dual model that includes both a familiarity preference and a preference for larger over smaller arrays. Furthermore, they make a strong empirical plausibility argument for their suggestion, for they show in an experiment quite similar (but not identical) to those in the literature that the longer looking at outcomes of 1 than at 2 in a 1 + 1 event and the longer looking times at outcomes of 2 than 1 in a 2 – 1 event are obtained even without the addition or subtraction part of the event, as predicted by the familiarity component of their dual model.

Although Cohen and Marks have made a strong plausibility argument for their alternative interpretation, they are far from showing that their two-factor model accounts for the extant addition/subtraction data. Most

[1] Wynn's experiment has been replicated and extended in many independent laboratories – with 4-month-old infants by Cohen and Marks, this volume; Simon, Hespos and Rochat, 1995; Koechlin, Dehaene and Mehlher, 1998; with older infants, Chiang and Wynn (2000), Huntley-Fenner, Carey and Solimando (under review), Uller *et al.* (1999), Feigenson, Carey and Spelke (in press); with non-human primates, Hauser, McNeilage and Ware, 1996; Uller, Hauser and Carey (in press). In spite of the failure to replicate by Wakeley, Rivera and Langer (2000), I do not take the replicability of these findings to be in doubt.

Address for correspondence: Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA; e-mail: scarey@wjh.harvard.edu

simply, they have not shown that the additions and subtractions make no difference in the standard paradigm, when there are only two outcomes (e.g. 1 + 1 = 2 or 1) rather than the four outcomes (e.g. 1 + 1 = 0 or 1 or 2 or 3) of their studies. It would be easy to directly test whether Cohen and Marks' model accounts for the extant data. The standard Wynn addition/subtraction experiment should be repeated under two conditions, exactly the same with respect to familiarization with the apparatus, and with respect to familiarization to the outcomes, if any. The only difference between the two conditions should be whether the addition/subtraction event occurs. If Cohen and Marks' model accounts for the standard data, then just as in their Experiments 1 and 3, the same pattern of data should obtain with and without the addition/subtraction event.

I doubt that this finding would obtain, for two reasons. First, I suspect that Cohen and Marks' data depend upon infants becoming overloaded by the complexity of the four-choice outcome. Second, and more importantly, there is abundant convergent evidence from other paradigms for number preserving representations of hidden sets (whether in the object file or analog magnitude formats), including evidence not subject to a familiarity preference reinterpretation.

But suppose that Cohen and Marks are right, and that the experiment proposed above shows that their dual factor model accounts for the extant infant addition/subtraction data? Where would this leave us with respect to theories of infant numerical representations? Evidence from the Wynn paradigm for infant representation of number would certainly have been undermined, for an alternative interpretation of the data from that paradigm would have been proposed and empirically supported. This would be a very important development; science advances by overturning accepted interpretations of reliable data. However, as I mentioned above, I believe this outcome unlikely, given *other* evidence for infant representation of number among infants this young or younger. And that other evidence would stand, for it is evidence not subject to the dual model reinterpretation.

To give one example of evidence that would stand: in the Spelke, Kestenbaum, Simons and Wein (1995) split screen study, 4-month-old infants always see only one object emerge from and return behind each of two separated screens. In one condition, the object appears in the middle, consistent with a single object going back and forth and in the other condition, no object appears in the middle, spatiotemporally specifying two numerically distinct objects, one behind each screen. After infants are habituated to this event, the screens are removed, revealing either one or two objects. The familiarity preference interpretation predicts longer looking at the single object outcome in both conditions, for infants have always only seen a single object. However, this result obtains only in the condition where the object never appears in the middle, the condition in which the one-object outcome is the unexpected outcome. Similarly, see Aguiar and Baillargeon's (1999) split screen studies with infants as young as 2.5 months of age, which also control for familiarity preferences. Young infants distinguish events with one or two objects in them and look longer at outcomes with the wrong number, not at the outcomes they have become familiarized to.

With older children (7 months and older), an avalanche of data support the existence of numerical representations, both in the object file (see Feigenson, Carey & Spelke, in press, and Feigenson, Carey & Hauser, in press, for reviews) and analog magnitude formats (Xu and Spelke, 2000). Some data do not depend upon looking times at all, so familiarity preferences are not an issue. For example, Feigenson, Carey and Hauser (in press) showed that 10- and 12-month-old infants update representations of hidden sets of crackers as additional ones are added, and make more/less choices over the represented sets. They choose the greater quantity when given a choice between 1 + 1 and 1, between 1 + 1 and 1 + a hand wave (controlling for time and attention drawn to the two buckets) and between 1 + 1 + 1 vs 1 + 1.[2]

Indeed, by the time infants are 7 or 8 months of age, there is good evidence that looking time patterns in the Wynn paradigm itself reflect attention drawn by violations of the expected number in the updated set rather than the Cohen and Marks dual model. Chiang and Wynn (2000) and Huntley-Fenner, Solimando and Carey (under review) have shown that success depends upon the nature of the individuals presented on the stage. Infants succeed if the entities are solid objects, and fail if they are piles of blocks or sand that are perceptually identical to the objects when at rest on the stage. In terms of Cohen and Marks' dual model, the familiarity preferences are identical in two conditions (object/piles) of these studies. Also, in Wynn and Chiang's (1998) 'magical appearance/disappearance' studies, infants looked longer at the unexpected outcomes only in the magical disappearance conditions, in spite of the fact that the number of objects the infants were familiarized with in the magical appearance and magical disappearance

[2] Object file representations, rather than analog magnitudes, most probably underlie infants' behavior in this task. Success or failure depends upon the size of the largest set, with a strict limit of 3. Infants succeed at choices of 1 vs 2 and 2 vs 3, but fail at 2 vs 4 and even 1 vs 4. Success is not a function of the ratio between the two sets, as dictated by the analog magnitude system of representation (Feigenson *et al.*, in press).

conditions were identical. These data violated both prongs of the dual model: the familiarity preference and the preference for more objects, for infants also did not look longer at outcomes of two objects than at outcomes of a single object.

In sum, there are data from several paradigms for numerical representations[3] in very young infants, evidence not subject to the dual model reinterpretation. This leads me to predict that the dual model will not turn out to be the right interpretation of the data from the standard Wynn paradigm, even from 4-month-olds. However, even if I am wrong in that prediction, data from slightly older infants, both within the Wynn paradigm and from other methods, supports the same representational capacities that the Wynn data have been taken to reveal. The most important issues, I would submit, concern the representational capacities of young infants and how these develop.

Cohen and Marks presuppose that Occam's razor favors resisting attributing 'sophisticated' abilities to young infants, and they believe that familiarity computations are less sophisticated than computations based on addition and subtraction over represented sets. This may be so, but unless Cohen and Marks spell out the nature of the representations underlying the familiarity judgment, we cannot tell. Familiarity requires a match between a stored representation and a representation of a currently perceived scene – what are those representations like, how are they constructed from perceptual input, and what is the computation that establishes the match/mismatch? Both the object file model and the analog magnitude model specify precisely the format of representation, the computations involved in updating models as changes are perceived and the nature of the comparison processes that determine familiarity (or novelty). Until Cohen and Marks provide a characterization of their dual model in such detail, we cannot meaningfully compare it to these other models with respect to sophistication.

The importance of specifying the format of infants' representations and the computations the infant can carry out over those representations is brought home when we consider Cohen and Marks' supposition that infants *learn* the effects of addition and subtraction sometime after 4 months of age. Infants are supposed to learn this through observation, as Mill (1973) and Kitcher (1985) propose. Perhaps, but how could such learning proceed without the antecedent ability to represent number? Fodor's famous dictum, 'one cannot learn what one cannot already represent' applies here. To learn the numerical consequences of addition and subtraction, one must have the capacity to represent the number in the initial set, the number in the addend or subtracted set, and the number in the final set. Both the analog magnitude model and the object file model assume that these consequences of adding and subtracting objects from sets do not have to be learned; they are part of the representational system from the beginning. But both specify the representations in such a way that it would be easy to model how the numerical consequences of addition and subtraction could be learned, just because both yield number relevant representations to enter into the learning procedure. The challenge for Cohen and Marks is to provide a learning procedure that does not require exactly the sophisticated representational and computational abilities they caution us about.

To conclude, I wish to emphasize my admiration for Cohen and Marks' elegant paper. Whatever final consensus emerges from ongoing research on the development of infant representations of number, Cohen and Marks' paper should certainly get the attention of the field concerning how important it is to control for familiarity preferences as we interpret data from the violation-of-expectancy looking time methods.

[3] One might question whether the object file model, if correct, is 'numerical'. I grant it is numerical in a much weaker sense than the analog magnitude model, for it contains no symbols for integers. However, it requires computations of numerical identity and computations of 1–1 correspondence, so pace Simon (1997), it is hardly 'non-numerical'.

References

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*, **39**, 116–157.

Chiang, W.C., & Wynn, K. (2000). Infants' representations and teaching of objects: implications from collections. *Cognition*, **77**, 169–195.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford: Oxford University Press.

Feigenson, L., Carey, S., & Hauser, M. (in press). Infants' spontaneous ordinal choices. *Psychological Science*.

Feigenson, L., Carey, S., & Spelke, E.S. (in press). Infants' discrimination of number vs continuous extent. *Cognitive Psychology*.

Hauser, M., McNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences*, **93**, 1514–1517.

Huntley-Fenner, G., Carey, S., & Salimando, A. (under review). Objects are individuals but stuff doesn't count: perceived rigidity and cohesiveness influence infants' representation of small numbers of discrete entities.

Kitcher, P. (1985). *The nature of mathematical knowledge*. Oxford: Oxford University Press.

Koechlin, E., Dehaene, S., & Mehler, J. (1998). Numerical transformations in five-month-old infants. *Mathematical Cognition*, **3**, 89–104.

Meck, W., & Church, R. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 320–334.

Mill, J.S. (1973). A system of logic ratiocinative and inductive (eighth edition). In J.M. Robson (Ed.), *Collected works of John Stuart Mill* (*Volumes 7 and 8*). Toronto: University of Toronto Press (original work published 1943).

Scholl, B.J., & Leslie, A.M. (1999). Explaining the infant's object concept: beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* (pp. 26–73). Oxford: Blackwell.

Simon, T.J. (1997). Reconceptualizing the origins of number knowledge: a 'non-numerical' account. *Cognitive Development*, **12**, 349–372.

Simon, T., Hespos, S., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, **10**, 253–269.

Spelke, E.S., Kestenbaum, R., Simons, D.J., & Wein, D. (1995). Spatio-temporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology*, **13**, 113–142.

Uller, C., Huntley-Fenner, G., Carey, S., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, **14**, 1–36.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, **358**, 749–750.

Wynn, K., & Chiang, W-C. (1998). Limits to infants' knowledge of objects: the case of magical appearance. *Psychological Science*, **9**, 448–455.

Xu, F., & Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, **74**, B1–B11.

Trying to build on shifting sand: commentary on Cohen and Marks

Kelly Mix

Department of Psychology, Indiana University, USA

Infant researchers have long relied on looking time procedures to gain access to the preverbal mind. Originally, these procedures were used to map out early perceptual abilities, examining, for example, changes in pattern perception (Fantz, 1958, 1961). Over the past several decades, however, there has been an explosion of research using infant looking times to tap increasingly sophisticated cognitive abilities, including numerical reasoning, physical reasoning and language processing. Indeed, infant looking times are the sole source of support for many extraordinary claims of early competence. Consider the following examples drawn from the number literature:

> I argue that human infants possess extensive numerical competence. Empirical findings show that young infants are able to represent and reason about numbers of things. Infants' ability to determine number is not based on perceptual properties of displays of different numbers of items, nor is it restricted to specific kinds of entities . . . [infants] possess procedures for operating over these representations in numerically meaningful ways, and so can appreciate the numerical relationships that hold between different numerical quantities. They thus can be said to possess a genuine system of numerical knowledge. These early capacities suggest the existence of an unlearned core of numerical competence. (Wynn, 1995, p. 35)

> . . . innate skeletal structures in the domain of number direct early attention to collections of separate entities (in any modality). These underlie infants' ability to match the numerosity of a visual display with the number of drumbeats they hear on a given trial . . . and keep track of the surreptitious effects of addition and subtraction. (Gelman, 1991, p. 313)

> . . . infants can relate the number of entities in one set to the number in another set, at least in regard to the equivalence or non-equivalence of the numerical magnitudes of the sets. They compute this relation even when the entities are objects and events that are presented in different modalities and bear no natural relation to one another . . . This finding suggests that infants are able to operate at a remarkably abstract level . . . (Starkey, Spelke & Gelman, 1990, p. 123)

Address for correspondence: Department of Psychology, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA; e-mail: kmix@indiana.edu

These interpretations are optimistic given the data on which they rest, perhaps recklessly so. This is because, despite their widespread use, surprisingly little attention has been paid to the processes that underlie looking time procedures themselves. In other words, as Marshall Haith (1980) put it, what are the rules that babies look by? Although a body of work produced over the past four decades provides some of the answers, this line of inquiry has taken a backseat to research that probes higher-level processing. Most researchers seem content to employ looking time measures without pursuing a deeper understanding of what makes them work.

Moreover, there is often little regard for the literature on this topic that already exists. Cohen and Marks (this issue) do the field a great service by reminding us of one well-documented 'rule' for infant looking – a preference for familiarity early in habituation. As they have elegantly demonstrated, such a preference, along with a tendency to look longer when more is presented, could account for Wynn's (1992) highly influential infant calculation finding. Their study is yet another piece of evidence that calls previous claims of sophisticated numerical reasoning in infants into question (see Mix, Huttenlocher & Levine, in press, for a review of others). It also adds to the research that has revealed similar confounds in other 'bedrocks' of infant cognition (e.g. Bogartz, Shinskey & Schilling, 2000; Cashon & Cohen, 2000). It is disturbing to contemplate how many other published studies could fall prey to the same criticism. Indeed, it seems that those of us trying to bridge infancy and early childhood risk building on a foundation of shifting sand.

One remedy to this situation is to learn more about the underlying processes that govern changes in infants' looking behavior. Until these processes are better understood in their own right, it is difficult to interpret any looking time experiment with confidence. This is true not only for the violation of expectation paradigm in which infants are familiarized rather than fully habituated, but also for habituation itself. Forthcoming work by Schoner and Thelen (2001) demonstrates that habituation and dishabituation are driven by a variety of parameters that usually have nothing to do with the central question being tested and are likely to be overlooked and unreported by most researchers. These include the perceptual strength of the stimuli (e.g. complexity), the length of time that elapses between trials, and individual differences in rates of habituation. When these parameters were varied in simulations using a dynamic field model of habituation, very distinct patterns of looking time emerged. Schoner and Thelen reached a sobering conclusion: 'The model showed that the balance between familiarity and novelty preferences can be changed by rather subtle differences in the stimuli or timing of events. If the parameters are not constrained by theory or mechanism, it is possible to adjust the stimulus displays to produce nearly any outcome one desires' (p. 38). And this is just what researchers inadvertently do when they tweak a procedure until they obtain a significant result.

Just as animal researchers are at risk for anthropomorphizing their non-human subjects, infant researchers are at risk for overlaying adult reasoning on basic perceptual responses. Longer looking times are commonly interpreted as evidence that infants have formed an abstract representation, compared it to a test stimulus, and effectively said to themselves, 'Hey, that's different!' or 'How surprising!' However, the tools used to probe infant cognition are governed by much more basic perceptual processes than those they are meant to tap. Cohen and Marks aptly point out that 'one should be cautious about attributing sophisticated cognitive processes to young infants when simpler processes will suffice' (p. 200). I couldn't agree more.

References

Bogartz, R., Shinskey, J.L., & Schilling, T.H. (2000). Object permanence in five-and-a-half-month-old infants? *Infancy*, **1**, 403–428.

Cashon, C.H., & Cohen, L.B. (2000). Eight-month-old infants' perception of possible and impossible events. *Infancy*, **1**, 429–446.

Fantz, R.L. (1958). Pattern vision in young infants. *Psychological Review*, **8**, 43–47.

Fantz, R.L. (1961). The origins of form perception. *Scientific American*, **204**, 66–72.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: initial and transcendent constructions. In S. Carey & R. Gelman (Eds.), *Epigenesis of mind: Essays on biology and cognition* (pp. 293–322). Hillsdale, NJ: Erlbaum.

Haith, M.M. (1980). *Rules that infants look by*. Hillsdale, NJ: Erlbaum.

Mix, K.S., Huttenlocher, J., & Levine, S.C. (in press). Multiple cues for quantification in infancy: is number one of them? *Psychological Bulletin*.

Schoner, G., & Thelen, E. (2001). A dynamic field model of infant habituation. Manuscript submitted for publication.

Starkey, P., Spelke, E.S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, **36**, 97–127.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, **358**, 749–750.

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, **1** (1), 35–60.

Do infants have numerical expectations or just perceptual preferences?

Karen Wynn

Department of Psychology, Yale University, New Haven, USA

Why do infants respond as they do in 'numerical reasoning' experiments? In Wynn (1992), when shown a '1 + 1' addition event, 5-month-old infants looked longer at incorrect outcomes of 1 and 3 than at the correct outcome of 2. When shown a '2 – 1' subtraction event, they looked longer at an incorrect outcome of 2 than at a correct outcome of 1. These results were interpreted within the framework of expectancy-violation: infants were computing the outcomes of these events, expecting a '2' outcome to a '1 + 1' event and a '1' outcome to a '2 – 1' event, and hence looking longer at the incorrect outcomes because they were unexpected.

Cohen and Marks present data from three experiments that, they suggest, support an alternative explanation for infants' performance. In their first experiment, 5-month-old infants saw either a series of '1 + 1' events or a series of '2 – 1' events, each with four different outcomes shown, of 0, 1, 2 or 3 objects. In both series, they showed longest looking, not to the three impossible outcomes over the one possible one, but to the outcome display identical to the initial display (i.e. to 1 object in the 1 + 1 series, to 2 objects in the 2 – 1 series) over the outcomes differing from the initial display. In the second experiment, when simply presented with displays containing 0, 1, 2 or 3 items, infants showed a preference for larger set sizes over smaller ones. In the third experiment, infants were presented with 'number-change' events (in which they were shown either 1 or 2 objects subsequently hidden by a screen; when the screen was removed, either 0, 1, 2 or 3 items were revealed); here, infants again looked longer at the (possible) outcome in which the number revealed was the same as that in the initial display, than they did when the number (impossibly) changed.

Cohen and Marks argue that these results suggest that infants in the original Wynn (1992) experiments were not showing preference (as evidenced by longer looking times) for the impossibility of the outcome, but for familiarity (as when, for example, 1 + 1 appeared to result in 1 object) or for a larger number (accounting for infants' longer looking to 3 over 2 following a '1 + 1' event). However, there are reasons to doubt this conclusion.

Cohen and Marks do not replicate the *findings* of Wynn (1992)

First, in Cohen and Marks' Experiment 1 (in which infants were presented with '1 + 1' and '2 – 1' events), they found the pattern of results obtained in Wynn (1992) – a pattern of longer looking to 1 over 2 in the addition group and to 2 over 1 in the subtraction group, *in the first block of test trials only*. In Wynn's experiments (and in replications of these experiments by other researchers: Koechlin, Dehaene & Mehler, 1997; Simon, Hespos & Rochat, 1995; Uller, Carey, Huntley-Fenner & Klatt, 1999), this pattern was obtained across all test blocks taken together as a whole, and was evident at the end of the experiment as well as at the beginning.

Second, in Wynn (1992), when shown a '1 + 1' addition that resulted in an outcome of either 2 or 3, infants looked significantly longer at the (impossible) outcome of 3 over 2; this preference was statistically significant across all test blocks taken together as a whole. But Cohen and Marks' group of '1 + 1' addition infants did not look significantly longer at an outcome of 3 over 2 at all, either in the first block, the second block, or across the two blocks taken together. In the first block, infants did look longer at 3 than at 2, but this was not significant. In the second block, infants showed the reverse (but again non-significant) pattern of preference, looking longer at 2 than at 3.

Cohen and Marks do not replicate the *methods* of Wynn (1992)

These failures to replicate Wynn's overall patterns of results suggest that Cohen and Marks' experiments are not tapping some of the cognitive processes operative in

Address for correspondence: Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520-8205, USA; e-mail: karen.wynn@yale.edu

Wynn's and others' experiments. This may be due to the fact that Cohen and Marks employed significant methodological deviations from the original Wynn (1992) study. In Wynn's experiments, infants were presented with just *two different outcomes* of the operation (either addition or subtraction) shown to them – and infants received *three repetitions of each outcome*, for a total of six test trials alternating between the correct and the incorrect outcome (in counterbalanced order). Cohen and Marks used a quite different design. In all of their experiments, they presented each infant with four distinct outcomes rather than two, each repeated twice, for a total of eight test trials.

We know that small physical alterations to an experiment can strongly affect infants' performance; for just one example, consider Hespos and Baillargeon's (2001) elegant and painstaking examination of infants' reasoning about occlusion versus containment. Young infants failed to make correct inferences about an object A that was placed inside another object B, but reasoned correctly about A when the very same object B served as an occluder rather than a container, by placing the to-be-hidden object *behind* it rather than *inside* it. Thus, even subtle changes can profoundly influence infants' performance. Cohen and Marks' significantly different methods, combined with their failure to replicate Wynn's original results, make comparison of their results with Wynn's difficult at best. It also weakens their argument: showing that processes (such as a preference for familiarity, preference for a larger number, etc.) are operative in one experimental situation is not sufficient as evidence that they are operative in other situations (e.g. the Wynn 1992 experiments). Their failure to replicate in itself shows that some processes are operative in the Wynn (1992) paradigm that are not being revealed in Cohen and Marks' paradigm.

Two speculations on the reason for the different results obtained by Cohen and Marks (I keep this brief as, given the absence of personal communications between Cohen and Marks and myself regarding our procedures, stimuli and set-up, there are no doubt many unknown differences between their experiments and mine, any of which may have contributed to the difference in results):

1. Providing 5-month-old infants with many different test outcomes (four in Cohen and Marks' experiment, as opposed to two in Wynn's original study), may increase infants' attention to the perceptual features of, and perceptual differences between, test trials – differences which are superficially evident and therefore easier to process – over the conceptual differences between them (detection of which requires inferential processes, which may be the first to suffer under conditions of information overload).
2. In a paradigm in which the majority of test trials depict impossible outcomes (75% of trials in Cohen and Marks' Experiments 1 and 3, as opposed to just half of trials in standard violation-of-expectation experiments such as Wynn's 1992 ones; that is, *triple the ratio* of impossible to possible events), infants may quickly learn to 'expect the unexpected', or that any outcome is 'possible' in this experimental context. This modification may therefore actually invalidate an expectancy-violation paradigm. (It would have been helpful if Cohen and Marks had data showing that some genuine violation of infants' expectations is detectable with this method.)

These two concerns raise the question, What would infants have done if shown, as in Wynn's experiments, only two different outcomes, one possible, one impossible? Using a between-subjects design, Cohen and Marks could have compared infants' looking to the same four outcomes of 0, 1, 2 and 3 objects. Such a design would have allowed more meaningful comparison of their results with those of the Wynn (1992) studies.

Cohen and Marks' data do not support their explanation

Cohen and Marks argue that in Wynn's (1992) study, infants looked longer at an outcome of 3 over an outcome of 2 to a '1 + 1' event due to a general preference for larger numbers, not due to having generated an expectation that the outcome should be 2. They offer as data supporting this argument the fact that, in their own Experiment 2 in which infants were simply presented with different numbers of objects to look at (without any prior operation rendering some of these 'possible', others 'impossible'), they found a trend for infants to give longer looks to larger set-size displays. However, this trend occurred in the second block of test trials only. *In their first block, infants actually looked longer at 2 than 3*. But in Wynn's experiment, infants showed a preference for 3 over 2 from the very beginning of the test trials. Moreover, to the extent that Cohen and Marks' infants looked longer at (albeit not significantly) a '1 + 1 = 3' outcome than a '1 + 1 = 2' outcome in their Experiment 1, they showed this preference in the *first* block of test trials only, not the second! Given Cohen and Marks' data that infants' preference for larger numbers does not emerge until *after* their first four test trials, such a preference cannot account for infants' non-significant preference for 3 in block 1 of Cohen and Marks' Experiment

1; nor can it account for infants' preference for 3 over 2 in the early trials of Wynn's study. In sum, Cohen and Marks' evidence actually suggests that to account for infants' preference for the impossible outcome of 3 over 2 in Wynn's (1992) study, one must appeal, *not to a preference for larger numbers, but to something else*; for one possibility, to an expectation on the part of infants for the correct *number* of objects.

Conclusions

To present a compelling alternative explanation for infants' performance, it is not enough to merely show that some processes that might, in theory, account for infants' performance (such as preference for familiarity or preference for larger sets over smaller) are operative at some points in some experimental conditions. One must show that such processes are operative at the *same* points and in the *same* conditions where infants' performance, the performance to be explained, is obtained. And, of course, the data presented in support of the alternative account should support the account, not undermine it. In both these respects, Cohen and Marks' alternative account is wanting.

References

Hespos, S.J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: a surprising discrepancy. *Psychological Science*, **12**, 140–147.

Koechlin, E., Dehaene, S., & Mehler, J. (1997). Numerical transformations in five-month-old human infants. *Mathematical Cognition*, **3**, 89–104.

Simon, T.J., Hespos, S.J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, **10**, 253–269.

Uller, C., Carey, S., Huntley-Fenner, G., & Klatt, L. (1999). What representations might underlie infant numerical knowledge? *Cognitive Development*, **14**, 1–36.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, **358**, 749–750.

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, **1**, 35–60.