Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Psychology 47 (2003) 367–401

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format

Marc D. Hauser* and Susan Carey

Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA

Accepted 21 January 2003

Abstract

The project of comparative cognition benefits from common measures across species. We report here on five experiments using the violation of expectancy looking time measure with free-ranging rhesus macaques (*Macaca mulatta*), each designed to build on current knowledge concerning spontaneous representations of number. Each subject, tested in only one experimental condition, watched as eggplants were placed behind a screen one at a time, after which the screen was removed revealing an outcome that either matched or did not match the number placed there. Subjects looked longer at impossible than possible outcomes in $1 + 1 = 2$ or 3, 1 small + 1 small = 1 big or 2 small, $2 + 1 = 2$ or 3, and $2 + 1 = 3$ or 4 conditions. They failed in $2 + 1 + 1 = 4$ or 3 or 5 and in $1 + 1 + 1 = 2$ or 3 conditions. This pattern of results closely matches that observed across several previous studies of human infants. The data allow us to test among four different proposals concerning the format and content of the mental representations underlying looking in these experiments. Object file representations are favored over: (i) low-level perceptual representations, (ii) representations of continuous variables such as volume or surface area, and (iii) analog magnitude representations of number. We conclude by considering exactly how the object tracking system revealed in these and other related experiments does and does not represent number, and how it might be one evolutionary precursor of the human specific system of number representations.

---

* Corresponding author. Fax: 1-617-496-7077.

E-mail address: mdhauser@wjh.harvard.edu (M.D. Hauser).

0010-0285/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0010-0285(03)00050-1

1. Introduction

A deep understanding of any human conceptual capacity requires at least an in principle understanding of its origin—both on evolutionary and ontogenetic time scales. Two goals of modern developmental cognitive science are to characterize the innate representational resources bequeathed by evolution to human beings and to characterize the learning/maturational mechanisms that bring the child from the initial state to an adult conceptual system. The first of these—the initial state—is our present concern. Comparative studies are essential to the enterprise of uncovering the evolutionary origins of human representational capacities (Hauser, 2000; Heyes & Huber, 2000). Evidence that some representational capacity is universal among primates, or present in species such as chimpanzees believed to be our closest living relatives, lends credence to the possibility that it is part of the innate endowment of human beings.

There are at least three crucial steps to establishing that a representational capacity is shared across species. First, one must show that members of the two different species have mental representations with the same content—e.g., both represent number, or artifacts, or spatial locations. This is not enough of course. Both bees and humans represent *pollen*, but we would not be tempted to say the two species share a system of representations. The human format of representation is often verbal and enters into both linguistic and theory-related computations; the bee representations are non-linguistic and atheoretical and enter into computations that support foraging and larva feeding. Second, to establish shared representational capacities, one must show that the format of representations are the same, as well as the computations carried out over them. And third, one should establish a common neural substrate or circuitry in order to assess whether the common computations are performed by homologous or analogous brain areas; similarity can arise, in evolution, either due to common descent (homologies) or to convergence (homoplasies). We focus here on the first two steps. We explore identity in conceptual content and format between representational systems deployed by human beings and non-human primates. We focus on prelinguistic human infants because it is likely that their mental representations are not formulated in a natural language format—infants do not think in French, or English, or Chinese. It is, however, an empirical question whether the representational systems of prelinguistic infants differ from those of non-human primates. Human infants will learn language, and the language acquisition device could, in principle, be manifest in infant representational capacities that non-human primates lack.

To maximize the interpretability of the cross-species comparison, the experiments should deploy the *same* methodological techniques. If the experiments with animals use extensive training procedures, so too should those with human beings. Conversely, if the experiments with humans recruit spontaneously available representational resources, so too should those with animals.

Here we report a series of experiments concerning primate (rhesus macaque) representation of objects and number. Many previous studies have shown that both human infants and non-human primates form representations of bounded, coherent, objects, tracing numerical identity of represented objects on the basis of spatio-temporal continuity (Munakata, Santos, O'Reilly, Hauser, & Spelke, 2000; Needham & Baillargeon, 1997; Simon, Hespos, & Rochat, 1995; Spelke, Kestenbaum, Simons, & Wein, 1995; Van de Walle, Carey, & Prevor, 2000; Whalen, Gallistel, & Gelman, 1999; Xu, 1999). Many studies also show that under some circumstances both human infants and non-human primates represent the number of objects in an array (Boysen, Berntson, Shreyer, & Quigley, 1993; Brannon & Terrace, 1998; Feigenson & Carey, in review; Matsuzawa, 1985; Olthof, Iden, & Roberts, 1997; Washburn & Rumbaugh, 1991; Xu & Spelke, 2000). That is, it has already been established that prelinguistic infants and non-linguistic non-human primates each have the capacity to form representations with the content *object* and *number.* For the most part, however, the experimental paradigms used to explore the representations of non-human primates and human infants differ greatly. Animal studies on number typically use methodologies that require extensive training, sometimes months or even years (Boysen et al., 1993; Brannon & Terrace, 1998; Matsuzawa, 1985). Obviously, such studies are not possible with human infants. The measures suitable for human infants, such as reaching or looking times in habituation/dishabituation paradigms or violation of expectancy techniques, tap spontaneous representations (Starkey, Spelke, & Gelman, 1990; Van de Walle et al., 2000; Wynn, 1992; Xu & Spelke, 2000). Typically, infants are shown events only once or just a few times, and the nature of their representations of the events is inferred from patterns of reaching or looking. Although the techniques used with human infants are driven by methodological necessity, spontaneous representations are of considerable theoretical interest in their own right, for they reflect what features of the world are salient to the participants in the studies, as well as providing evidence concerning their natural, untutored, representations of the world.

Given the theoretical interest of spontaneous representations, as well as the importance of common methodologies across species, it is of considerable importance that the methods used with human infants also yield interpretable data with non-human primates, and with other more distantly related species as well. To date, we have used three different paradigms to study spontaneous representations of both infants and non-human primates: a two-box spontaneous choice paradigm (Feigenson, Carey, & Hauser, 2002a; Hauser, Carey, & Hauser, 2000; Sulkowski & Hauser, 2000), a manual search paradigm (Santos, Sulkowski, Spaepen, & Hauser, 2002; Van de Walle et al., 2000), and a violation of expectancy looking time paradigm (Feigenson, Carey, & Spelke, 2002b; Hauser, MacNeilage, & Ware, 1996; Uller, Hauser, & Carey, 2001; Wynn, 1992). In addition to establishing common methods across nonverbal members of both species, these studies have yielded remarkable convergences in patterns of data. For example, and of direct relevance to the current studies, the two-box choice paradigm reveals the set size signature of object file representations both in rhesus macaques (Hauser et al., 2000) and in 10- and 12-month-old human infants (Feigenson et al., 2002a, 2002b). In this task, participants watch as pieces of food are placed, one at a time, into each of two closed containers, and are then allowed to approach one of the them. Participants succeed (pick the larger number) if both sets are small (upper limit 4 for rhesus adults and upper limit 3 for babies) but fail if one of the sets exceeds that limit (e.g., rhesus fail at 3 vs. 8; infants fail at 2 vs. 4, 3 vs. 6, and even 1 vs. 4).

The present studies extend our comparative research program on spontaneous representations of number in primates focusing in particular on the violation of expectancy looking time method originally developed by Wynn (1992). Hauser et al. (1996) showed that this method yields interpretable data with free-ranging rhesus macaques, who look longer at impossible than at possible outcomes in $1 + 1 = 1$ vs. 2 comparisons, as well as in $2 - 1 = 2$ vs. 1 comparisons; the convergence in results across species occurred even though Hauser et al. used a between-subjects design whereas studies of infants use a within-subjects design. Uller et al. (2001) found that laboratory reared cotton-top tamarins, tested on a within-subject design, also succeed in $1 + 1 = 2$ or 1 experiments, consistently looking longer at the impossible outcome of 1 than at the possible outcome of 2. One goal of the present series of studies is to extend this work and to bring data to bear on constraining the format and content of the representations underlying rhesus performance in this paradigm. A secondary goal of these experiments is to assess whether the looking time method provides convergent or divergent results when contrasted with the two-box choice paradigm successfully implemented with both rhesus (Hauser et al., 2000) and human infants (Feigenson et al., 2002a).

In the infant case, the content and format of representations underlying looking in the violation of expectancy looking time methods in these "number" experiments is currently under debate. There are four distinct proposals that have current adherents. At issue is the nature of the representation of the events before the outcome is revealed, the nature of the representation of the outcome, the basis of comparison between the two, and the source of relative attention (longer looking). At one extreme is the proposal of Cohen and Marks (in press), who have argued that for 4-month-old infants, at least, neither object nor number representations are implicated. Rather, they suggest that the set-up events and the outcomes are specified in terms of a low-level perceptual similarity space of shape, color, texture, etc. On their model, attention is a function of degree of perceptual match between the most frequently seen stimulus during the events before the outcome is revealed and the outcome itself. They explain the longer looking at the familiar match, rather than at the novel match, in terms of the complexity of these events (Hunter & Ames, 1988). In most violation of expectancy looking time experiments with infants, the most familiarized stimulus array happens to be the unexpected outcome. Consider, for example, Wynn's $1 + 1 = 2$ or 1 experiment. Infants see an object resting on the stage floor, the screen is raised, and then a second object is added. The unexpected outcome of 1 object on the stage floor matches the array infants have seen. In a clever experiment supporting this hypothesis, Cohen and Marks showed that the same pattern of looking time is observed even if the second object is not added. That is, given a $1 = 1$ or 2 comparison, infants look longer at 1, just as they do in a $1 + 1 = 1$ or 2 comparison. Similarly, in a $2 = 2$ or 1 comparison, infants look longer at 2, just as they do in a $2 - 1 = 2$ or 1 comparison. This result is consistent with Cohen and Marks' hypothesis, which we will call the "*low-level perceptual familiarity hypothesis*."

In the case of adult rhesus macaques, details of the Hauser et al. (1996) procedure allow us to rule out Cohen and Marks' low-level perceptual familiarity hypothesis. Hauser and colleagues used a between-subjects design; each monkey saw two familiarization events and just one test event, which was an impossible outcome for one group of monkeys and a possible outcome for the other group. Crucially, in each group, monkeys were familiarized with the outcome of their test event during both familiarization events, and thus those in the possible and impossible outcome groups were equally familiar with the outcomes. Thus, the difference between the groups could not be a function of a familiarity preference (or a simple novelty preference) formulated over a perceptual similarity space. Because the design employed by Hauser and colleagues provides one way to rule out the low-level perceptual familiarity hypothesis, and provides robust results under natural conditions with rhesus monkeys, we adopt this technique for all of the experiments presented in this paper.

A second proposal also denies that Wynn-like experiments reflect representations of number. This proposal does not, however, deny that infants and animals are creating summary representations of all of the objects placed behind the screen, including those added and subtracted. Instead, the proposal is that infants are encoding these objects in terms of some continuous variable such as total contour length, total front surface area, or total volume, and not number. Two series of studies (Clearfield & Mix, 1999; Feigenson et al., 2002a) have shown that representations of such continuous variables sometimes underlie looking times in simple "number" habituation/dishabituation paradigms. Feigenson et al. (2002a) also present data consistent with the proposal that 7-month-old infants are representing total front surface area or total volume in $1 + 1 = 2$ or 1 and $2 - 1 = 2$ or 1 violation of expectancy experiments. For example, in a 1 small + 1 small = 1 big (expected front surface area and volume, unexpected number) vs. 2 big (unexpected front surface area and volume, expected number) paradigm, infants looked longer at the unexpected total surface area/volume, and generalized familiarization when the number was unexpected but the total front surface area/volume was expected. We shall call this the "*continuous quantity hypothesis*."

In a possible processing model that implements the continuous quantity hypothesis, total front surface area or total volume or total contour length is represented as an analog magnitude. Take front surface area as an example. In the Wynn paradigm, infants accumulate and store a representation of total front surface area as objects are added and subtracted from the array, thus creating a representation of the total expected in the outcome array. Upon seeing the outcome array they create a representation of the total front surface area of the array, and compare the two representations as they would compare any represented continuous variable such as length, brightness, distance or time. The hypothesis that infants are comparing their representations of the arrays behind the screen with their representations of the outcome array on the basis of some continuous variable is consistent with the data from Feigenson et al. (2002a). Two sources of data, however, suggest that these are not the *only* representations underlying infant and monkey looking times in these experiments. Uller (1997) and Uller et al. (2001) showed that 8-month-old infants and cotton-top tamarins looked longer at the outcome that was a numerical mismatch in a 1 small + 1 small = 2 small vs. 1 big comparison, where the single large outcome matched the expected outcome in both total volume and total front surface area. Experiment 2 explores this issue in the case of free-ranging rhesus macaques.

A third proposal, suggested by several authors, claims that symbolic *number* representations, most likely formulated over analog magnitude representation of number, underlie infant and monkey performance in the small number addition/subtraction experiments (Dehaene, 1997; Gallistel & Gelman, 2000; Wynn, 1998). The idea is that the numerical value of the set-up displays is represented by an analog magnitude, akin to a number line. This analog magnitude might be created by either an iterative (Meck & Church, 1983) or parallel (Church & Boradbent, 1990; Dehaene, 1997) mechanism. Gallistel and Gelman (2000) have championed the iterative proposal of Meck and Church, for this mechanism implements a counting algorithm and could underlie the child's learning to count. On this proposal, a fixed amount of energy is let through to an accumulator for each individual in an array, irrespective of its size, such that the total energy accumulated is a linear function of number. It is this value that is stored in short term memory. The numerical value of the outcome display is then represented by the same system, and the two values compared. A mismatch draws more attention than a match. We call this proposal the "*number analog magnitude hypothesis*." This hypothesis differs from that above in that the content of the analog magnitude in this case is number, whereas in the former case it is a continuous variable such as total front surface area.

The fourth and final proposal denies that analog magnitude representations of number are recruited in the violation of expectancy looking time methods with small numbers (Leslie, Xu, Tremoulet, & Scholl, 1998; Simon et al., 1995; Simon, 1999; Uller, Carey, Huntley-Fenner, & Klatt, 1999). Rather, each object in the array is represented by a unique symbol. Building on work in mid-level vision (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn & Storm, 1998; Trick & Pylyshyn, 1994), the symbol for each individual has been called an "object file." We shall call this proposal the "*object file hypothesis*." The resulting representation *implicitly* represents the number of objects in the array, as there is only one open object file for each object in the set, and object files are opened or closed as objects are added or subtracted from the hidden set. In test trials, the representation of the hidden set is compared to the representation of the revealed set on the basis of 1–1 correspondence.

Two types of data have been brought to bear on the choice between number analog magnitude and object file models of performance on any given task. First, performance breaks down with different number comparisons in the two cases. Object file representations are subject to a set size limit—only small sets can be individuated in parallel and stored in short term memory. In contrast, large sets can be encoded by analog magnitudes; discrimination is subject to Weber's law, such that sets of a constant ratio may be discriminated. Thus, there is a set size signature of object file representations (performance is subject to a set size limitation and not to the ratio between sets) and a Weber-fraction signature of analog magnitude representations (performance is limited by the ratio between sets). In the two-box choice procedure, performance of both monkeys and infants shows the set size signature of object file representations rather than the Weber-fraction signature of analog magnitude representations (Hauser et al., 2000; Feigenson et al., 2002a, 2002b). Second, object file representations are transparently subject to variables that affect short term memory for multiple items. In contrast, it is not yet clear whether such variables influence analog magnitude representations in which a single value is held in short term memory. For example, Uller et al. showed that young infants in the looking time studies of small number addition were more likely to fail if they were required to make multiple updates in working memory prior to assessing the correctness of test outcome, a result that is not predicted on the basis of the iterative analog magnitude models. That is, infants succeed in "object first" paradigms (Wynn, 1992) a full 6 months earlier than they succeed in "screen first" paradigms (Uller et al., 1999). In object first tasks, the infant sees the first object on the stage floor before a screen is introduced and a second object placed behind it. In "screen first" paradigms, in contrast, the screen is introduced first, and objects placed one at a time behind it. The latter task thus requires two successive updates in working memory.[1] Uller and colleagues therefore argued that the results favored object file representations. In the present studies we ask whether monkeys' representations are affected by the number of updates in working memory that the task requires. We turn to the question of how this sensitivity bears on model choice in the general discussion.

Previous studies with non-human animals, including primates, have established that species including pigeons, parrots, rats, squirrel monkeys, rhesus macaques, and chimpanzees create analog magnitude representations of number (Biro & Matsuzawa, 1999; Boysen & Bernston, 1989; Boysen et al., 1993; Brannon & Terrace, 1998; Davis & Perusse, 1988; Fernandes & Church, 1982; Kawai & Matsuzawa, 2000; Matsuzawa, 1985; Murofushi, 1997; Olthof et al., 1997; Pepperberg, 1994; Rilling & McDiarmid, 1965; Roberts, Coughlin, & Roberts, 2000; Rumbaugh & Washburn, 1993; Thomas & Lorden, 1993; Washburn & Rumbaugh, 1991). These experiments tap large numbers and performance is limited by the ratios of the sets to be discriminated. However, without exception these experiments require extensive training, sometimes months of daily training, and are not appropriate methods for studies with infants.

The present experiments have two goals. First, they constitute a continued exploration of the viability of violation of expectancy looking time methods for studying the spontaneous mental representations of rhesus monkeys in the wild. Second, and foremost, we seek data that further constrain the format and content of the representations that underlie patterns of looking by rhesus macaques in the Wynn paradigm. We seek convergent evidence to that from the choice method that the spontaneous representations of small sets of objects by monkeys and prelinguistic humans are the same, both in content and format, and that object file representations underlie monkeys' performance.

To date, only two numerical operations ($1 + 1 = 2$ vs. $1 + 1 = 1$; $2 - 1 = 1$ vs. $2 - 1 = 2$) have been studied with rhesus macaques using the looking time method

---

[1] Notice that both versions of this task require the infant to successively attend to each object as it is introduced onto the stage. Certainly, incrementing an accumulator requires attention to each individual, but the two versions place equivalent demands on attention.

(Hauser et al., 1996). Thus, the format and even the content of the representations that underlie successful discrimination in this case is still an entirely open issue. With respect to content, we explore whether it is number that is being tracked, as opposed to low-level perceptual features of the arrays or some continuous variable correlated with number. Even if it is number that is being tracked, it is not even clear that exact numerosity is represented, rather than just that $1 + 1$ is different from 1, or more than 1. Experiment 1 begins with the question of exact numerosity and Experiment 2 begins with the contrast between number and continuous variables. All of the studies in this series bear on Cohen and Marks' low-level perceptual familiarity hypothesis. Experiments 3–5 explore the upper limits on monkeys' performance and the costs of multiple updates on encoding of the sets of objects.

2. General methods

2.1. Subjects

All experiments were conducted on a population of rhesus monkeys (*Macaca mulatta*) living on the island of Cayo Santiago, Puerto Rico. Rhesus monkeys were introduced to the island in 1938 and the population has been under intensive investigation ever since (Rawlins & Kessler, 1987). Studies on Cayo Santiago have focused on demographic changes in population structure, the causes and consequences of different behavioral patterns, vocal communication, and cognitive ability (Bercovitch & Berard, 1993; Gouzoules, Gouzoules, & Marler, 1984; Hauser, 1998, 2001; Rendall, Rodman, & Edmond, 1996). As a result, much is known about rhesus monkeys, and in particular, the population on Cayo Santiago.

There are approximately 1000 rhesus monkeys on Cayo Santiago. The population is divided into seven social groups. Individuals can be identified by chest and leg tattoos, as well as distinctive ear notches. In this population, females reach reproductive maturity at approximately 3 years, whereas males reach reproductive maturity at approximately 4 years. Subjects for the following experiments were adult males and adult females.

2.2. Procedural synopsis

The logic underlying each of the experiments was the same. Specifically, we used the expectancy violation procedure to assess whether rhesus monkeys look longer at impossible than at possible events. As in other experiments using this procedure, an event is defined as impossible if the outcome following a series of events with objects is inconsistent with what human adults know about general physical principles. In each experiment, outcome (possible/impossible) was a between-subjects factor. Each monkey received two familiarization trials, designed to introduce it to the apparatus, the screen, the objects, the events, and the outcomes. In the familiarization events, there were no impossible outcomes. In most of the experiments, in both possible and impossible conditions, the outcomes during familiarization were exactly the same as those during test. Thus, differential looking at possible and impossible outcomes cannot be a function of novelty or familiarity preferences for arrays presented during familiarization.

Each experiment proceeded as follows. A team of three researchers set out to find either a lone individual or an individual sitting with a small group. We sought subjects that were resting, rather than foraging, grooming, fighting or mating. One of the three researchers was responsible for presenting the objects and events to the subject, one was responsible for filming the subject's looking time onto video (JVC-D1 digital camera), and one was responsible for recording the test conditions, especially information on problems associated with a trial (e.g., subject distracted by other individuals; experimenter error during the display).

The experimenter responsible for presenting each trial set up 3–5 m away from the test subject. The display box was made out of foamcore and measured 91 cm × 62 cm × 46 cm. The front screen slid up and down a track so that it could be easily put into place or removed. On the back of the screen was a cloth pouch. Impossible events were produced in one of two ways. First, for outcomes in which one fewer object than expected was revealed on the stage, one of the objects was placed into the pouch, thereby giving the illusion that an object had been added to the array on the stage floor. For example, in Experiment 3 ($2+1=2$ vs. $2+1=3$), we created the violation ($2+1=2$) by placing the third object into the pouch; when the screen was removed, this third object failed to appear, leaving only the two original objects. Second, we created a trap door in the back wall of the display box so that we could add or change objects, in order to produce impossible outcomes with extra objects on the stage floor or with different objects on the stage floor.

During both familiarization and test events, the experimenter only made eye contact with the subject when presenting objects and either placing or removing the screen; when the looking time period commenced, the experimenter stared straight down until the 10 s period ended. It was necessary for the experimenter to look at the subject during the presentation phase to make sure that the subject was looking at the relevant display events. We excluded subjects from the final analysis if they: (1) failed to look at any portion of the display event prior to timing the duration of their looks and (2) were distracted by another individual during the presentation or moved away before the test trial. We did not keep a record of why individuals were excluded, but from previous experiments we can estimate that approximately 10–15% of those excluded failed to look during the events as objects were placed on the stage and the rest were excluded for being distracted by another individual or for moving away prior to the test trial. Although these strict criteria for inclusion caused us to eliminate between 30 and 50% of subjects per condition, our final data set included only those animals that watched all aspects of each display of each trial.

We videotaped all trials and then digitally acquired (30 frames/s) them onto a Macintosh G3/G4 with Adobe Premiere. We then scored each trial blind to condition by limiting the video footage to the subject's face and having each trial scored by two or more individuals who were ignorant of the test conditions run; inter-observer reliability scores ranged from 0.89 to 0.99. These methods have been reliably used in several other studies of rhesus monkeys (Hauser et al., 1996; Munakata et al., 2000).

3. Experiment 1: 1 + 1 = 2 vs. 3

Experiment 1 is an extension of the original violation of expectancy looking time experiments on Cayo Santiago (Hauser et al., 1996) in which rhesus monkeys looked longer at an impossible outcome of $1+1=1$ than at a possible outcome of $1+1=2$. In Experiment 1 we explore whether rhesus monkeys expect precisely 2 objects following the presentation of a $1+1$ operation, or whether they simply expected that the outcome of a $1+1$ event would be different from 1 or more than 1. Following Wynn (1992) we compared $1+1=2$ test events (possible) with $1+1=3$ test events (impossible). If rhesus merely expect that $1+1$ should be different from or more than 1, they should not differentiate between these two test events. However, if they expect $1+1$ to be precisely 2, they should look longer at the impossible outcome of 3 objects.

3.1. Methods

A total of 68 rhesus monkeys were tested. Of these individuals, we were able to complete testing of 14 subjects run on $1+1=2$ (possible condition) and 16 subjects run on $1+1=3$ (impossible condition).

Fig. 1A diagrams the events in the possible and impossible conditions of Experiment 1. Each session consisted of two familiarization events and a single test event. The familiarization events in this experiment served to introduce the subject to the presentation of a stage, to the objects (i.e., eggplants) that would be seen during the test events, and to the actions associated with introducing eggplants behind the barriers. None of the outcomes during familiarization were impossible. Furthermore, the subjects in each condition were equally familiarized to the outcomes of the test events.

Possible condition (1 + 1 = 2). This is a direct replication of Hauser et al. (1996). The stage was introduced with two eggplants resting on the stage floor. As soon as the stage was in position, we started the 10 s looking time period. This first familiarization (F1) was then followed by a second (F2) in which the stage was presented with occluder in place. We then placed one eggplant behind the occluder, followed by the addition of a second eggplant, with each eggplant placed on opposite sides of the display box; once an eggplant was placed behind the screen, we withdrew our hand empty. The occluder was then removed revealing two eggplants, and a 10 s looking time period started. In the test condition, we presented an empty stage, set up the occluder, and then placed one eggplant followed by a second eggplant behind the occluder. The occluder was then removed, revealing two eggplants; we then timed a 10 s looking period.

Notice that in the second familiarization trial, the outcome of two objects was possible but not necessary. Because the trial began with the occluder in place, the

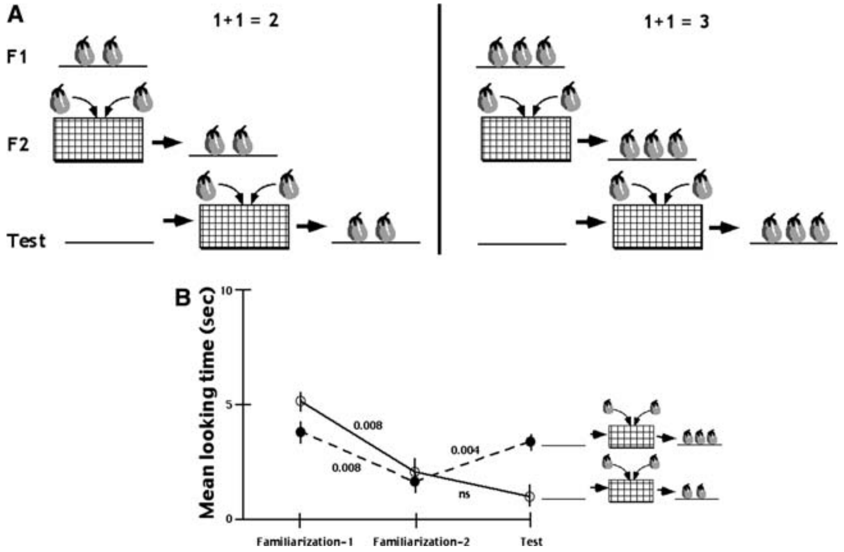

Fig. 1. (A) Familiarization (F1 and F2) and Test trials for possible ($1 + 1 = 2$) and impossible ($1 + 1 = 3$) conditions. (B) Results from both conditions. Mean (= /− SE) looking time (s) is plotted on the y-axis, and the presentation trials on the x-axis.

monkeys could not know for sure how many objects to expect behind the stage. In the test trial, however, the outcome of two eggplants is not only possible, but necessary (barring a magic trick); the monkeys see the empty stage before the occluder is put into place.

Impossible condition ($1 + 1 = 3$). The unoccluded stage was introduced with three eggplants resting on its floor. As soon as the stage was in position, we started the 10 s looking time period. This first familiarization trial (F1) was then followed by a second (F2) in which the stage was presented with occluder in place. One eggplant was then placed behind the occluder, and the hand removed empty. A second eggplant was then placed behind the occluder, and the hand removed empty; the eggplants were placed on opposite sides of the display box. The occluder was then removed revealing three eggplants, and a 10 s looking time period started. Notice that because F2 starts with the occluder in place, the outcome of three eggplants is possible (i.e., subjects should not have any expectations about the presence or absence of eggplants behind the occluder). In the test condition, we presented an empty stage, put the occluder in place, and then placed one eggplant followed by a second behind the occluder. The occluder was then removed, revealing three eggplants; a 10 s looking time period was then started. Because we surreptitiously placed a third eggplant through the trap door in the impossible test condition, the same action was mimicked in the possible test conditions so that all actions were standardized; the only difference between these two tests was the actual outcome when the occluder was removed.

In this and all subsequent experiments, the results will be analyzed in two complementary ways. First, we discuss the habituation across the two familiarization trials, and the recovery from habituation as a function of whether the test trial was a possible or impossible event. Success consists of continued habituation to the possible test event in the face of significant recovery of interest in the case of the impossible test event. Second, we compare the looking times in the two test events (possible/impossible) directly. Successful discrimination consists in longer looking at the impossible outcome than at the possible outcome. Throughout the presentation of results, reported probability values are two-tailed.

3.2. Results

Fig. 1B shows the mean looking time data for subjects tested in the possible and impossible conditions of Experiment 1. For the first familiarization, subjects looked slightly longer at a display of two eggplants then at a display of three eggplants, but this difference was not statistically significant ($t = 1.75, p = .09$). From the first to the second familiarization, subjects in both groups showed a statistically significant decline in looking time (in each case, $p < .01$); there was no difference in looking time between groups in the second familiarization. From the second familiarization to the test, subjects in the possible condition showed a non-significant decline in looking time ($t = 2.91$, $p = .10$), while subjects in the impossible condition showed a significant increase in looking time ($t = 9.81, p = .004$). Thus, the monkeys generalized habituation between the familiarization trials and the test trial in the possible condition, but recovered interest, relative to the second familiarization trial, to the test event of the impossible event.

Analyzing the test events alone, looking times to the impossible outcomes were reliably longer than looking times to the possible outcomes ($t = 3.98$, $p = .004$).

3.3. Discussion

Experiment 1 tested the hypothesis that the success in the Hauser et al. (1996) $1 + 1 = 2$ vs. 1 experiments merely reflected the expectation that the outcome should be different from or more than 1. Both outcomes in the present experiment, 2 eggplants and 3 eggplants, are more than and different from 1 eggplant, yet the monkeys clearly differentiated the possible and the impossible outcomes. The monkeys' attention was significantly increased only to the impossible outcomes. The difference in looking time to 2 vs. 3 eggplants in the test trials cannot be accounted for by an overall preference for 3 because there was no such preference during the familiarization trials. The difference cannot be due to a preference for more objects emerging over the course of the experiment, because no such pattern was observed in the Hauser et al. (1996) $1 + 1 = 2$ vs. 1 experiments, which used an identical design to this one. And finally, the difference cannot be due to a familiarity preference because the monkeys in both groups were equally familiarized to the outcomes they saw in the test events. The monkeys in the possible condition had seen outcomes of 2 objects, the outcome of their test trial, during each of the two familiarization trials and the monkeys in the impossible condition had seen outcomes of 3 objects, the outcome of their test trial, in each of the two familiarization trials. These results, along with those of Hauser et al. (1996) rule out Cohen and Mark's (in press) low-level perceptual familiarity account of monkeys' success in this task, according to which the preference for the impossible outcome in the test trial is due to a familiarity match in terms of perceptual features of the arrays that are presented during familiarization.

4. Experiment 2: 1 small + 1 small = 2 small vs. 1 Big One

We conclude from Experiment 1 that adult rhesus monkeys are capable of forming representations of the exact number of objects behind the screens, updating these as objects are added to or subtracted from the arrays, at least with sets of one, two, and three objects. These experiments do not, however, reveal how rhesus monkeys represent these arrays, nor do they reveal the computations underlying the mismatch between the representations of the hidden set and the revealed set in the impossible outcomes. As Feigenson et al. (2002b) point out, the impossible outcomes in all of the violation of expectancy number experiments with monkeys and human babies violate the expected values of continuous variables (volume/surface area) as well as the expected number of objects. Feigenson et al. disentangled these variables in experiments with 7-month-old human infants, providing a $1 + 1 = 1$ big vs. 2 big comparison, where each big object in the outcome had twice the front surface area and volume as each object that had been placed on the stage. Thus, the outcome consisting of one big object was the expected amount of material but the unexpected number, and the outcome of two big objects was the expected number but an unexpected amount of material. Seven-month-old infants looked longer at the outcomes that violated the expected amount of material, not the expected number. This experiment confirms the suggestion emerging from simple habituation paradigms (Clearfield & Mix, 1999; Feigenson et al., 2002b) that infants find differences in continuous variables (total volume, total surface area, and total contour length) more salient than differences in number in these looking time paradigms. However, Uller (1997) and Uller et al. (2001) demonstrated that in experiments with human infants and cotton-top tamarins, using a within-subjects design, it is not the case that subjects *only* track total volume or surface area. When shown 1 small object added to an array containing 1 small object, both human 8-month-olds and tamarins look longer at outcomes of a single large object, twice the volume of each those in the array, than at the outcome of 2 small objects, in spite of the fact that the total volume is the same in the two outcomes. Furthermore, in Hauser et al. (2000) box search paradigm, rhesus monkeys preferentially selected a box with 3 pieces of apple over a box with 1, even though the overall volume of these two quantities was equated.

Experiment 2 explores whether rhesus monkeys are encoding *only* total expected volume in the context of a looking time procedure. When rhesus monkeys see purple eggplants disappear behind an occluder, they might maintain a representation of the volume of purple stuff rather than the number of purple eggplants. When the occluder reveals either one or three eggplants, this represents a violation of the volume of purple stuff anticipated. If so, an outcome consisting of one large eggplant, twice the volume of each of the small eggplants introduced behind the occluder in an $1 + 1$ event, would not be unexpected.

4.1. Methods

We tested 44 adult rhesus monkeys on a $1 + 1 = 2$ vs. 1 Big One condition. Of those tested, our final data set included 7 individuals tested on $1 + 1 = 2$ (possible) and 10 individuals tested on $1 + 1 = 1$ Big One (impossible). We have labeled the Big One outcome as impossible because two small eggplants can not turn into a large one, even though the large eggplant does preserve the original volume or amount of purple stuff. The procedure is diagrammed in Fig. 2A.

Possible condition (1 small + 1 small = 2 small). Subjects tested on the possible condition were given the same familiarization and test trials as subjects tested on the possible condition of Experiment 1. This condition is thus a direct replication of the possible condition of Experiment 1, as well as of the possible condition of the second experiment in Hauser et al. (1996).

Impossible condition (1 small + 1 small = 1 big). The first familiarization of the impossible condition (F1) involved the presentation of a single, large eggplant equal in volume to the two small eggplants presented in the possible condition; a 10 s looking time period started as soon as the display box was in position. The second familiarization (F2) involved the placement of two eggplants, sequentially, behind the occluder; following the placement of each eggplant, the hand was removed empty. The occluder was then removed, and two eggplants revealed. In the test trial, we first

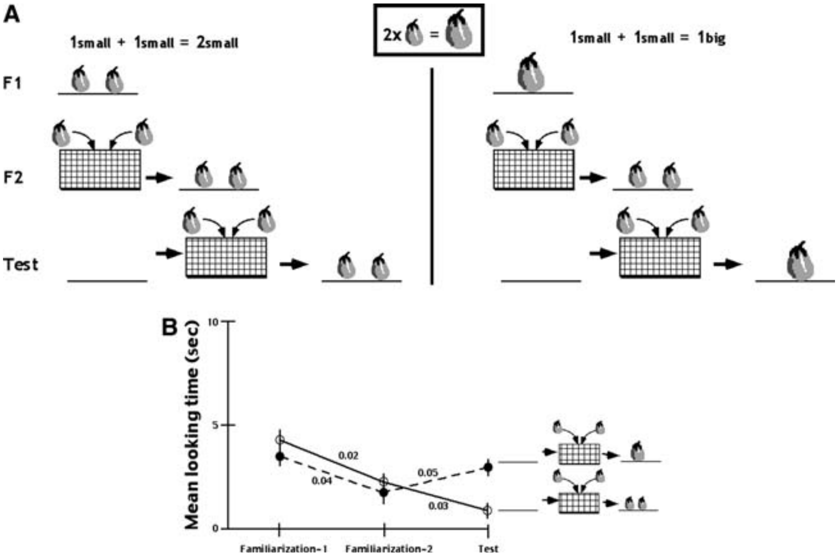

Fig. 2. (A) Familiarization (F1 and F2) and Test trials for possible ($1 + 1 = 2$) and impossible ($1 + 1 =$ Big One) conditions. (B) Results from both conditions. Mean (= /– SE) looking time (s) is plotted on the *y*-axis, and the presentation trials on the *x*-axis.

presented an empty stage. The occluder was then put into place, followed by the sequential placement of two eggplants. When the occluder was removed, a large eggplant was revealed. The illusion was carried out by placing the two small eggplants into the pouch behind the occluder, and then placing the large eggplant through the trap door. As in Experiment 1, we controlled for the actions carried out in the impossible test condition by reaching through the trap door during the possible test condition; by repeating the same actions for both test trials, differences in looking time between conditions can only be accounted for by differences in the object or objects presented once the occluder has been removed.

In this experiment, it was not possible to carry out the full design of Experiment 1, in which the first and second familiarization events each involved the same outcome as in the test events. In the second familiarization trial of the impossible condition, we could not show the same outcome as in the test trial, because this outcome would have been impossible. Nonetheless, if the monkeys were encoding *only* the total amount of stuff, the familiarization trials in both the possible and impossible conditions were all identical; they differed only in the spatial arrangement of the total volume of eggplant.

4.2. Results

The results are shown in Fig. 2B. For the first familiarization, there were no statistically significant differences in looking time between groups run on the possible and impossible conditions ($t = 1.15$, $p = .11$). Subjects in both groups showed a statistically significant decline in looking time from the first to the second familiarization ($p < .04$, t tests, in each case). Subjects in the possible condition showed a statistically significant decline in looking time from the second familiarization to the test ($t = 2.50$, $p < .03$). Subjects tested in the impossible condition showed a statistically significant increase in looking time from the second familiarization to the test ($t = 2.00$, $p < .05$). Thus, subjects in the possible condition (a replication of the $1 + 1 = 2$ condition of Experiment 1) generalized habituation from the second familiarization to the test event, whereas those in the impossible condition recovered interest to the outcome of the test event.

Subjects also looked longer in the impossible than the possible test ($t = 3.14$, $p < .02$), showing that these two outcomes were clearly discriminated.

4.3. Discussion

If subjects in Experiment 2 were tracking only volume or amount of purple stuff, they should not have differentiated the two test outcomes. It is unlikely that the longer looking at the 1 large object outcome was due to the subjects in the impossible outcome condition being less familiarized with the test outcome than the subjects in the possible outcome condition. Dishabituation to the impossible outcome in Experiment 2 was comparable to that in Experiment 1, even though in Experiment 1 the monkeys had been familiarized with the outcome of the test trial twice. Also, even though the monkeys in the possible outcome conditions saw the test outcome twice during familiarization, compared to only once in the possible outcome condition, Uller et al. (2001) found comparable results in an experiment with cotton-top tamarins. The Uller et al. experiment used a different familiarization procedure, and a within-subject design for possible/impossible outcomes. Specifically, although subjects were familiarized equally to the possible and the impossible outcomes during familiarization, they nonetheless looked longer at the impossible outcome of 1 large object than at the expected outcome of 2 small objects each the size of the objects they saw placed behind the stage. Uller (1997) also found the same pattern of results with 8-month-old human infants.

That the monkeys did differentiate the possible and impossible outcomes is consistent with two different interpretations. The first possibility is that they differentiated the outcomes on the basis of a mismatch in the number of objects in the display, the impossible outcome of 1 object not matching their representation of 2 objects involved in the events. Second, their representations of the eggplants behind the stage may have included a specification of the size of each eggplant they had seen placed there, and the single large eggplant in the impossible outcome differed from each of those. That is, their attention may have been drawn by a mismatch at the level of a single eggplant rather than by a mismatch in the number in the set. This experiment cannot decide between these two interpretations. However, Experiment 2 rules out the hypothesis that overall spatial extent, amount of stuff, or volume alone drives the differential looking times to possible and impossible outcomes in these studies.

5. Experiment 3: $2+1=3$ vs. 2 vs. 4

If either object file representations or analog magnitude representations of number underlie the monkeys' performance on this task, they should succeed with larger contrasts than 2 vs. 1. Evidence with respect to the question of the upper limit of infant performance is mixed. Wynn (1995) showed that 5-month-old infants succeeded at a $3-1=2$ vs. $2+1=2$ comparison, suggesting that they can represent the distinction between 3 and 2 under these circumstances. Uller (1999), in contrast, showed that 10-month-olds succeeded in a $2-1=2$ vs. $2-0$ (empty hand) $=2$ contrast, but not a $3-1=3$ vs. a $3-0$ (empty hand) contrast. To date, the issue of the upper limits on monkey performance in this paradigm has not been addressed. Experiments 3 and 4 begin to do so.

Given that monkeys succeed robustly in a 2 vs. 3 and a 3 vs. 4 comparison in the box choice addition experiments (Hauser et al., 2000), as well as a 2 vs. 3 comparison in the box choice subtraction experiments (Sulkowski & Hauser, 2000), we expected monkeys to succeed in a $2+1=3$ condition (possible outcome) relative to both $2+1=2$ (impossible outcome) and $2+1=4$ (impossible outcome) conditions. Experiment 3 tested both of these contrasts.

Experiments 1 and 2, like those in Hauser et al. (1996), were screen-first designs. In a screen-first design the monkey is first shown an empty stage, the screen is introduced, and all of the objects are introduced onto a hidden stage. Uller et al. (1999)

found that infants succeed at screen-first $1 + 1 = 2$ or 1 experiments a full 6-months later than they succeed at the more common object-first design, in which the first object or objects are presented visibly on the stage floor, then the screen is raised, and additional objects are added or removed. Because we were exploring larger numbers for the first time in Experiment 3, we adopted the easier object-first design.

5.1. Methods

A total of 101 adult rhesus monkeys were tested. The final data set included tests of 15 individuals on a $2 + 1 = 3$ condition (possible) and 16 individuals each on a $2 + 1 = 2$ and a $2 + 1 = 4$ condition (both impossible). Fig. 3A shows the design of these three conditions.

Possible condition ($2 + 1 = 3$). In the first familiarization trial of the possible condition, subjects were presented with a stage with 3 eggplants resting on its floor. In the second familiarization, the stage was presented with occluder in place. One eggplant was added behind the occluder, the experimenter's hand exited empty, and then the occluder was removed to reveal 3 eggplants. Note that this is a possible outcome because the stage was introduced with the occluder in place, thereby precluding any reasonable expectations about the number of eggplants in place behind the occluder. In the test trial, 2 eggplants were placed on the empty stage while the monkey watched and then the occluder was put into position. One eggplant was then placed behind the occluder, and the experimenter's hand was retracted empty. The occluder was removed, revealing 3 eggplants.

Impossible condition ($2 + 1 = 2$). In the first familiarization trial, the stage was introduced with 2 eggplants resting on its floor. In the second familiarization, the stage was presented with the occluder in place. One eggplant was added behind the occluder, the experimenter's hand exited empty, and then the occluder was removed revealing 2 eggplants. In the test trial, 2 eggplants were placed on the stage while the monkey watched and then the occluder put into position. One eggplant was then placed behind the occluder, and the experimenter's hand was retracted empty. The occluder was removed, revealing 2 eggplants.

Impossible condition ($2 + 1 = 4$). In the first familiarization trial of the $2 + 1 = 4$ condition, the stage was introduced with 4 eggplants resting on the floor. In the second familiarization, the stage was presented with occluder in place. One eggplant was added behind the occluder, the experimenter's hand was retracted empty, and then the occluder removed revealing 4 eggplants. In the test trial, 2 eggplants were placed on the stage while the monkey watched and then the occluder put into position. One eggplant was then placed behind the occluder, and the experimenter's hand removed empty. The occluder was removed, revealing 4 eggplants.

5.2. Results

Fig. 3B presents the results from both the possible and impossible conditions. In the first familiarization, there were no statistically significant differences in looking time between groups ($p > .05$); thus, subjects had no preference for larger over smaller sets

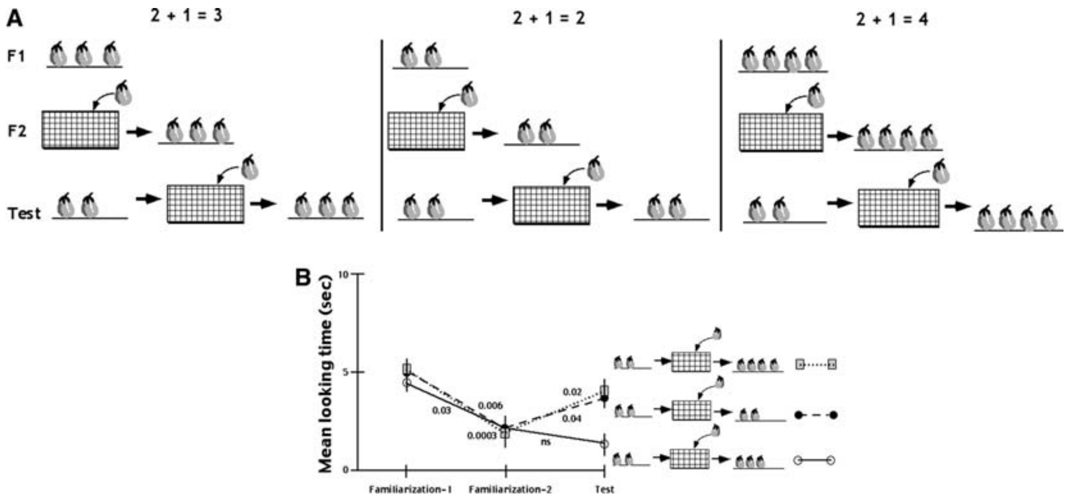

Fig. 3. (A) Familiarization (F1 and F2) and Test trials for possible (2 + 1 = 3) and impossible (2 + 1 = 2 and 2 + 1 = 4) conditions. (B) Results from all three conditions. Mean (=/− SE) looking time (s) is plotted on the *y*-axis, and the presentation trials on the *x*-axis.

of eggplants. Subjects in all three conditions showed a statistically significant decline in looking time from the first to the second familiarization trials ($.0003 < p < .03$, t tests). Subjects tested in the possible condition showed a decline in looking time from the second familiarization to the test trial, but this decrease was not statistically significant ($t = 0.95$, $p = .41$). In contrast, subjects tested in each of the impossible conditions showed a statistically significant increase in looking time from the second familiarization to the test ($2 + 1 = 2$; $t = 1.85$, $p < .04$, and $2 + 1 = 4$, $t = 6.50$, $p < .02$).

The differentiation of possible and impossible outcomes on the test trials was also assessed directly. Monkeys looked longer at the impossible outcome of 2 eggplants than at the possible outcome of 3 eggplants ($t = 7.13$, $p < .001$) and also the impossible outcome of 4 eggplants than at the possible outcome of 3 eggplants ($t = 7.42$, $p < .001$). There was no difference between the looking times at the two impossible outcomes.

5.3. Discussion

Subjects showed no preference for any of the outcomes during familiarization and subjects in all three groups habituated between the first and second familiarizations to the same extent. Thus, the success in this task cannot be due to a familiarity preference, as the outcomes of the test trials were equally familiar in all three conditions. Nor can the success be due to a preference for larger sets emerging over trials, for monkeys looked equally long at impossible outcomes whether they were smaller than the possible outcome or larger than the possible outcome, paralleling the results in the $1 + 1$ experiments (Hauser et al., 1996, and Experiment 1). We conclude that when rhesus monkeys see 1 object added to 2 others that are out of sight behind an occluder, they update their representation of the original 2 such that they expect precisely 3 eggplants, not 2 or 4.

6. Experiment 4: 2 + 1 + 1 = 3 vs. 4 vs. 5

In Experiment 3, rhesus discriminated 2 vs. 3 and 3 vs. 4. Experiment 4 continues to explore the upper limit on monkeys' success at this task, by including a 4 vs. 5 contrast. Experiment 4 also begins to explore the effects of manipulations of information processing demands within the 3 vs. 4 comparison that monkeys succeed on in Experiment 3. Experiment 4 sets up three conditions: $2 + 1 + 1 = 4$ (possible), $2 + 1 + 1 = 3$ (impossible), and $2 + 1 + 1 = 5$ (impossible). Given the failure in the 4 vs. 5 condition of the Hauser et al. (2000) two-box choice experiments, we expected the monkeys to fail to discriminate outcomes of 4 from outcomes of 5. It is the comparison of possible outcome of 4 with the impossible outcome of 3 that is of particular theoretical interest. Note that this design maintains a 3 vs. 4 comparison, as in Experiment 3, where $2 + 1 = 3$ was compared with $2 + 1 = 4$. One difference between Experiments 3 and 4 is whether the larger number is in the outcome array or the set up array. Second, building the representation of the hidden set during the test trial of Experiment 4 requires two updates in memory rather than one. That is, first a representation of 2 objects must be updated when another is added and then the resulting representation must be updated again. As Uller et al. (1999) found with human infants, conditions with two updates are more difficult than conditions with one. Experiment 4 asks whether these manipulations affect the monkeys' success on this task, just as they do infants'.

6.1. Methods

A total of 86 adult rhesus monkeys were tested. The final data set included 17 subjects tested on $2+1+1=3$ (impossible), 14 subjects tested on $2+1+1=4$ (possible), and 14 subjects tested on $2+1+1=5$ (impossible). Fig. 4A diagrams the procedure of Experiment 4.

Possible condition (2 + 1 +1 = 4). For all three groups, the first familiarization (F1) began with the introduction of a stage with 2 eggplants already resting on its floor, followed by a 10 s looking time period. For the group tested with an outcome of 4 eggplants (possible outcome), the second familiarization (F2) involved the presentation of a stage with occluder in place. Subsequently, 2 eggplants were sequentially placed behind the occluder, the occluder removed, and 4 eggplants revealed; looking time was scored for 10 s as soon as the occluder was removed. In the test trial, we presented 2 eggplants on stage, followed by the placement of an occluder. Next, 2 eggplants were sequentially placed behind the occluder, the occluder removed, and 4 eggplants revealed.

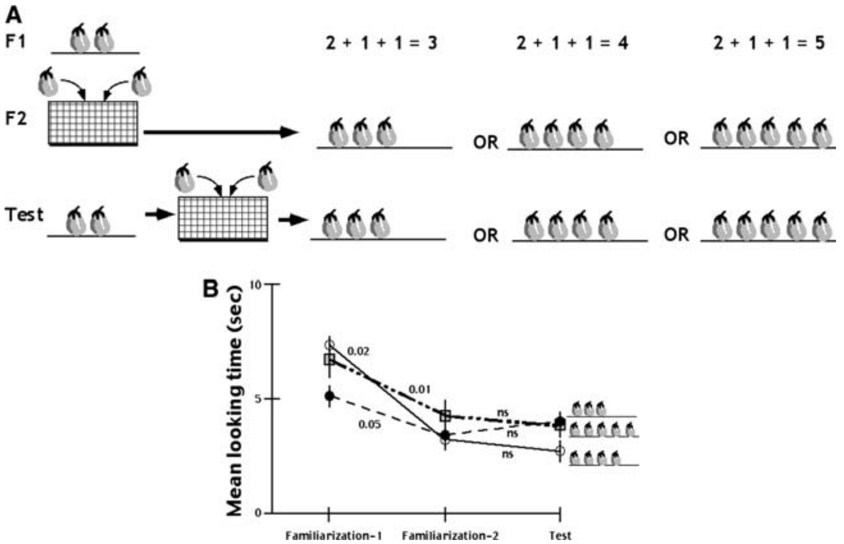

Fig. 4. (A) Familiarization (F1 and F2) and Test trials for possible $(2+1+1=4)$ and impossible $(2+1+1=3,\ 2+1+1=5)$ conditions. (B) Results from all three conditions. Mean (=/− SE) looking time (s) is plotted on the *y*-axis, and the presentation trials on the *x*-axis.

Impossible condition $(2+1+1=3)$. For the group tested with an outcome of 3 eggplants (impossible outcome), the second familiarization (F2) began with the presentation of a stage with the occluder in place. Subsequently, we placed 2 eggplants, sequentially, behind the occluder and then removed the occluder to reveal 3 eggplants; we scored looking time for 10 s as soon as the occluder was removed. In the test trial, we first presented 2 eggplants on stage, followed by the placement of an occluder. Next, 2 eggplants were sequentially placed behind the occluder, the occluder removed, and 3 eggplants revealed.

Impossible condition $(2+1+1=5)$. For the group tested with an outcome of five eggplants (impossible outcome), the second familiarization (F2) involved the presentation of a stage with occluder in place. Subsequently, 2 eggplants were sequentially placed behind the occluder, the occluder removed, and 5 eggplants revealed; we scored looking time for 10 s as soon as the occluder was removed. In the test trial, two eggplants were presented on stage, followed by the placement of an occluder. Next, 2 eggplants were sequentially placed behind the occluder, the occluder removed, and 5 eggplants revealed.

Note that in this design the first familiarization for all groups involved two eggplants (the first step in all test trials as well) and the second familiarization involved the same outcome that would be shown on the test. Thus, the monkeys in all three conditions were equally familiarized with the outcome of the test trial in their own condition to equal degrees. We mention this here because it is directly relevant to Cohen and Marks' low-level perceptual hypothesis.

6.2. Results

The results are shown in Fig. 4B. On the first familiarization, there was no difference in looking time between subjects presented with 4 as opposed to 5 eggplants, and no difference in looking time between subjects presented with 3 as opposed to 5 eggplants ($p > .05$) However, subjects looked significantly longer at an outcome of 4 eggplants than at an outcome of 3 eggplants ($t = 2.02$, $p < .05$). Subjects in all three groups showed a statistically significant decline in looking time from the first to the second familiarization trial ($.01 < p < .05$, t tests). There was no difference between groups in looking time for the second familiarization trial ($p > .05$). Subjects in all three groups failed to show a statistically significant change in looking time from the second familiarization to the test trial ($p > .05$).

Comparing the test trials directly, there were no differences in looking times to either of the impossible outcomes when compared to the looking times to the possible outcome of four eggplants ($p > .05$, t tests).

A final analysis compared looking times in the impossible outcome of 3 eggplants in Experiment 4 $(2+1+1)$ with the impossible outcome of 4 eggplants in Experiment 3 $(2+1)$. The groups differed significantly ($t = 3.13$, $p < .02$), confirming success at detecting the anomaly in Experiment 3 in the face of failure in Experiment 4, in spite of the same numerical comparisons (3 vs. 4) being involved.

6.3. Discussion

Subjects failed to detect a violation when the result of a $2 + 1 + 1$ operation resulted in outcomes of either 3 objects or 5 objects. That is, subjects looked as long to a possible outcome of 4 as they did to either of the impossible outcomes of 3 and 5. The failure to distinguish outcomes of 4 vs. 5 is consistent both with the object-file hypothesis (assuming an upper limit of 4) and with the number analog magnitude hypothesis (assuming that the contrast between 4 and 5 falls short of the ratio that can be discriminated). More importantly, the monkeys succeeded in a $2 + 1 = 3$ or 4 comparison in Experiment 3 and failed in a $2 + 1 + 1 = 3$ or 4 comparison in Experiment 4. As far as the Weber fraction comparison of the outcomes, the two experiments are identical. Two differences between the experiments may have contributed to the greater difficulty of Experiment 4. First, in Experiment 4, it is the representations of the larger of the two sets (4) that must be computed during the set-up event and held in working memory, whereas in Experiment 3 it is the smaller (3). Second, in Experiment 4, the representation of the original set seen on the stage floor must be updated twice during the set-up event, whereas in Experiment 3 it must be updated only once.

7. Experiment 5: 1 + 1 + 1 = 2 vs. 3

Experiments 3 and 4 differed not only in the number of updates required of the monkeys, but also in the size of the set that must be held in memory before the outcome was revealed. Experiment 5, which was designed on the basis of a finding with human infants, tests whether number of updates *alone* affects performance. Baillargeon, Miller, and Constantine (1994) found that 10-month-old infants had a fragile success (success on the first pair of test trials alone) in a $2 + 1 = 2$ or 3 experiment, but failed outright in a $1 + 1 + 1 = 2$ or 3 experiment. In this comparison, the only difference between the two conditions is the number of updates of the representation of the set of objects when hidden.

In Experiment 3 monkeys successfully discriminated the possible outcome in a $2 + 1 = 3$ condition from the impossible outcome of a $2 + 1 = 2$ condition. Experiment 5 tests exactly the same comparison of outcomes, except that the initial event in each case was $1 + 1 + 1$. In Experiment 3, monkeys saw the two objects on the stage, a screen was placed in front of them, and the monkeys had to update only once a representation initially formed from perceptual information alone; thus, Experiment 3 was an object-first design. Experiment 5, in contrast, is a screen first paradigm. The screen covers an empty stage, and three objects are introduced, one at a time. Three updates of the representation formed from perception (the empty stage) are required. The size of the resultant set (3) is identical in the two experiments, and the size of the set in the impossible outcome (2) is similarly identical. The only difference between the two experiments, then, was the number of times the monkey had to update a representation of the hidden objects behind the screen as new objects were added.

7.1. Methods

A total of 47 adult rhesus monkeys were tested on a 1 + 1 + 1 operation. The final data set included 17 subjects tested on the possible condition of 1 + 1 + 1 = 3 and 15 subjects tested on the impossible condition of 1 + 1 + 1 = 2. Fig. 5A diagrams the procedure.

Possible condition (1 + 1 + 1 = 3). In the first familiarization (F1) of the possible condition, we presented subjects with 3 eggplants already visible on the stage floor; we then scored a 10 s looking time period as soon as the stage was in position. In the second familiarization (F2), we presented subjects with the stage and occluder in place, then sequentially placed 2 eggplants behind the occluder, removed the occluder to reveal 3 eggplants, and then scored looking time for 10 s. This outcome is possible because subjects never see what is behind the occluder before we add the eggplants. In the test trial, we first present an empty stage and then put the occluder in place. Three eggplants are then placed in sequence behind the occluder. When the occluder is removed, subjects see 3 eggplants on the stage; looking time is scored as soon as we remove the occluder.

Impossible condition (1 + 1 + 1 = 2). In the first familiarization (F1) of the impossible condition, we present subjects with 2 eggplants already visible on the stage floor and score a 10 s looking time period as soon as the stage is in position. In the second familiarization, we present subjects with the stage and occluder in place, sequentially lower 2 eggplants behind the occluder, remove the occluder to reveal 2 eggplants, and then score looking time for 10 s. In the test trial, we first present an empty stage

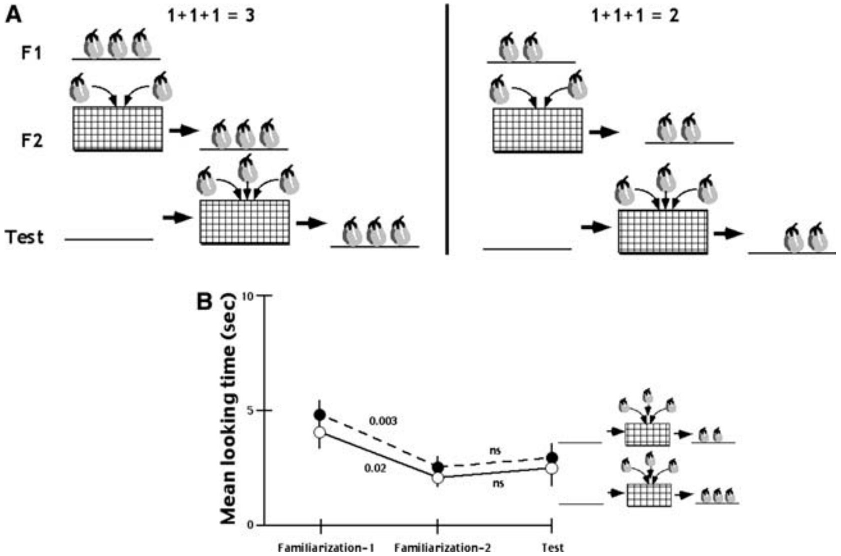

Fig. 5. (A) Familiarization (F1 and F2) and Test trials for possible (1 + 1 + 1 = 3) and impossible (1 + 1 + 1 = 2) conditions. (B) Results from each condition. Mean (= /− SE) looking time (s) is plotted on the *y*-axis, and the presentation trials on the *x*-axis.

and then put the occluder in place. We then place 3 eggplants in sequence behind the occluder, remove the occluder to reveal 2 eggplants on the stage, and then score looking time for 10 s.

7.2. Results

Fig. 5B presents the results from Experiment 5. Subjects showed no preference for looking at two as opposed to three eggplants in the first familiarization ($t = 0.84$, $p = .41$). Subjects in both groups showed a statistically significant decline in looking time from the first to the second familiarization ($.003 < p < .02$). Subjects in both groups then showed a statistically non-significant increase in looking time from the second familiarization to the test trial ($p > .05$).

A direct comparison between the looking times in the test events of the possible and impossible conditions revealed no difference ($p > .05$, t tests).

Finally, the impossible outcome in Experiment 5 was directly compared to the impossible outcome in Experiment 3. Monkeys looked significantly longer in Experiment 3 ($t = 2.13$, $p < .05$), reflecting a violation of expectancy in the object first $2 + 1 = 2$ condition in the face of failure to detect the anomalous outcome in the screen first $1 + 1 + 1 = 2$ condition.

Because of the importance of this result to our argument, we sought to replicate it to be certain of its robustness. We ran an exact replication of Experiment 5, with 12 monkeys in the possible condition and 11 monkeys in the impossible condition. Just as in Experiment 5, the monkeys again failed. In neither group did they recover interest from the second familiarization, and looking times at the impossible outcomes (mean = 2.58 s) did not differ from looking times at the possible outcomes (mean = 2.84 s; $t = 0.32$, $p = .75$).

7.3. Discussion

Results from Experiment 3 suggest that when rhesus monkeys see 1 object added to 2 originally visible but then occluded objects, they expect 3 objects, not 2. However, when the same outcomes are presented, but the operation involves three addition (+1) operations, rhesus fail to discriminate the possible and impossible outcomes. Again, rhesus macaques are sensitive to the same variables that affect success by human infants on this task (Baillargeon et al., 1994). It appears that for human infants and adult rhesus monkeys, repeatedly updating the representation causes problems for the mechanism that computes a representation of the hidden set. We turn next to a more general discussion of our results, and the implications they have for deciding between the four proposed models of number representation, including an assessment of content and format.

8. General discussion

The experiments presented in this paper had two aims. First, we sought to provide further evidence that the violation of expectancy method provides robust data in studies of non-human animals. Second, we sought to refine our understanding of the content and format of numerical representations in non-human primates, specifically rhesus monkeys. Concerning the first aim, our results provide clear evidence that the violation of expectancy looking time method yields highly reliable and interpretable data with non-human primates, even under free-ranging conditions. Replicating the design and results of Hauser et al. (1996), the monkeys in these experiments generalized from familiarization to test in a $1 + 1 = 2$ condition (possible outcome of Experiments 1 and 2), while showing a recovery of interest from familiarization to test in the impossible outcomes ($1 + 1 = 3$, Experiment 1, and $1 + 1 =$ big one, Experiment 2). The monkeys also showed generalization from familiarization to test in the $2 + 1 = 3$ condition, while showing a recovery of interest from familiarization to test in the impossible outcomes of $2 + 1 = 2$ and $2 + 1 = 4$ (Experiment 3). In light of these successes, the failures in Experiments 4 and 5 ($2 + 1 + 1 = 4$ or 3 and $1 + 1 + 1 = 2$ or 3, respectively) are not uninterpretable negative findings, but rather, interpretable findings derived from the same procedures, with results bearing directly on the format of representation guiding the monkeys' attention in this paradigm; this point bears on the second aim of our paper.

Experiments 1 and 3, along with those in Hauser et al. (1996), rule out Cohen and Marks' (in press) low-level perceptual familiarity account of monkeys' performance on this task. Because monkeys in both the possible and impossible conditions are familiarized only with the outcomes they actually see in the test trials, and because familiarization rates (declines in looking times between the first and second familiarization trials) are identical between conditions of each experiment, it is reasonable to assume that the monkeys have equally encoded the perceptual features of the arrays on the stage floor. Differential looking at the impossible outcomes cannot, therefore, be due to simple familiarity (or novelty) computed over perceptual features of the arrays seen during familiarization. Rather, the monkeys must be updating the representation of the empty stage (Experiment 1) or the stage with 2 eggplants on it (Experiment 3) as new eggplants are introduced behind the screen. We therefore conclude that the low-level perceptual familiarity hypothesis can not account for the rhesus monkey's performance in these experiments and thus, a different mechanism must mediate their attentional resources. Note also that the low-level perceptual similarity preference hypothesis cannot account for the results in the box search or two-box choice paradigms at which both human infants and rhesus macaques succeed, for in these paradigms there is no outcome revealed that could be the basis of a perceptual match.

Monkeys responded to the impossibility of the big eggplant outcome when they had been shown 2 eggplants each one half of its volume introduced behind the screen (Experiment 2). This result rules out the proposal that the match between the representation of the occluded objects and the outcomes in the test trials is based on volume or spatial extent alone. We conclude that some spontaneous representation of number controls attention in this task. That the monkeys succeeded in the $1 + 1 = 2$ vs. $1 + 1 = 3$ conditions of Experiment 1 shows that the representations of number are exact; it is not the case that the monkeys expect additions merely to increase the number in the original set. Finally, the failure of the monkeys in a

4 vs. 5 comparison in Experiment 4 is consistent with both object file representations (because 5 exceeds the limit on parallel individuation of object files, at least as established in humans and presumed to hold for non-human animals) and analog magnitude representations (on the assumption that 4:5 exceeds the Weber ratio monkeys can discriminate under these circumstances). This failure converges with the 4 vs. 5 failure in the box choice paradigm (Hauser et al., 2000), as well as in recent studies of tamarins using auditory stimuli (Hauser et al., in review).

Experiments 3 through 5, taken together, show that rhesus are sensitive to information processing demands of these tasks in a manner that closely parallels infants' sensitivity. Furthermore, we suggest that there are several reasons why details of this sensitivity support the object file model over the iterative analog magnitude model of Meck and Church. First of all, consider the success in Experiment 3 (set-up of 3, choice of 3 or 4) compared to the failure in Experiment 4 (set-up of 4, choice of 3 or 4). Both of these comparisons are 3 vs. 4, and thus should be equally difficult on the analog magnitude model. But on the object file model, the set-up must be held in short term memory as a file of symbols, one for each object in the array, and the outcome compared to it on the basis of 1–1 correspondence. Holding a larger set in memory (4 vs. 3) is more difficult. And if 1–1 correspondence must be computed when evaluating the outcome displays, then the comparison of 4 with 4 (the possible outcome in Experiment 4) is more difficult than the comparison of 3 with 3 (the possible outcome in Experiment 3).

Second, consider the sensitivity to number of updates during the set-up event. According to the iterative analog magnitude model, a subject increments an accumulator a fixed amount for each individual (object or event) introduced behind the screen, increments a different accumulator a fixed amount for each individual revealed in the test outcome, and compares the two on the basis of the accumulated magnitudes. As far as incrementing the accumulator for each additional object, it should not matter (at least based on current descriptions of this model) whether the objects are screened or not. But on the object file model, the participant creates a representation of the object or objects on the stage floor and must update that representation by adding an additional object when one is introduced. Each update in working memory requires changing the resultant model.

Of course a different version of the analog magnitude model, in which analog magnitudes are not computed until *after* a mental model of the objects on the stage floor has been created in short term memory, is consistent with the data presented here. On this alternative, *both* object-file representations and analog magnitude representations would be implicated in these studies. This is possible, of course, but as of yet there is no positive evidence that analog magnitudes are implicated in these small number comparisons.

A supporter of the analog magnitude account of the representations might argue that attention to each individual is needed to increment the accumulator, and the multiple update experiments provide more opportunity for attention to wander, and thus, for noise to enter the computation. Remember that monkeys were included in the sample only if they attended to the entire sequence of events. Nonetheless, we agree that this is a reasonable reply in the case of the studies reported here (for example, 1 + 1 + 1 of Experiment 5 compared to 2 + 1 of Experiment 3). However this reply does not apply to the Uller et al. comparison between the object first and the screen first conditions. In both of these cases, the infant had to attend to just two objects, introduced one at a time (1 + 1). The only difference was whether the screen was introduced before either of the objects was placed on the stage or after one had been placed on the stage floor. Infants failed in the screen first condition until 10-months of age, succeeding in the object-first condition at 4 months. Our interpretation of this finding, like others presented here, is that there is a cost to updating a working memory representation consisting of object files.

The argument in favor of object-file representations does not rest only on the pattern of sensitivity to number of updates in working memory. A much stronger argument derives from the set-size signature of object files. In the two-box forced choice task (Hauser et al., 2000), monkeys watched while different numbers of apple slices were placed, one at a time, into each of two opaque boxes. The dependent measure was the first box approached, followed by the retrieval of its contents. Each monkey received only 1 trial, eliminating the possibility of training; like the present studies, this paradigm taps spontaneous representations of small sets of object. The monkeys selected 2 over 1, 3 over 2, 4 over 3, and 5 over 3, but showed no preference for 4 vs. 5, 5 vs. 6, 4 vs. 8 or even 3 vs. 8. Moreover, these results cannot be due to a mechanism that taps timing of number of actions, as a control condition ruled out these possibilities. Specifically, when the number of objects and amount of time was equated by adding a rock to the smaller apple quantity, subjects continued to select the box with more apple slices, up to 4 over 3. These results are not consistent with the analog magnitude system of representation, as success is not a function of the ratio of the two numbers. The choice of 4 over 3 is a less favorable ratio than 8 vs. 4 or 8 vs. 3, yet monkeys succeeded in the former case and failed in the latter. The monkeys' performance showed the set-size signature of the object file model; as long as one of the choices is over 5, the monkeys perform at chance.

Feigenson et al. (2002a) showed that this measure of spontaneous representations yields interpretable data from preverbal human infants (10- and 12-month-olds). Critically, the infants' behavior also exhibits the set-size signature of the object file system, though in contrast with rhesus, infant performance falls apart at 4 rather than 5. When infants watch as an experimenter places different numbers of crackers into one of two opaque boxes, they choose the larger number for 1 vs. 2 and 2 vs. 3, but fail at 3 vs. 4, 2 vs. 4, and 3 vs. 6 comparisons. These data not only provide evidence for the object file system, but like the current studies, show that the same methods can be used across these two subject populations, providing closely parallel patterns of results.

The representations that underlie performance in the expectancy violation looking time task are spontaneously available. Monkeys are not trained; indeed, each monkey sees only three events. They are not reinforced for attending to the objects, for individuating them, for tracking numerical identity during occlusion, or for representing the exact number of objects in the set that results from adding an additional object to a hidden set of one or two. In these respects, the experiments presented here and in Hauser et al. (1996) are identical to the violation of expectancy looking time experiments carried out with human infants, in spite of many methodological differences driven by the contrast in testing conditions. These methodological differences bear further study, including attempts to reduce the differences or explain their contribution to the patterns of results. Because the monkeys are free to move away (and frequently do so as evidenced by the proportion of subjects excluded from the final analysis), the present experiments give only one test trial, and contrast possible and impossible outcomes between-subjects rather than the within-subjects designs favored for infants. Time spent looking in the monkey experiments, in contrast to infant studies, are not participant controlled; each trial has a fixed 10 s recording period. The experimenter is in view in the monkey experiments but not in the infant studies. Nonetheless, the comparisons between Experiment 3, on the one hand, and Experiments 4 and 5, on the other, reveal a pattern of findings that is identical to that observed with human 8- to 10-month-old infants (Baillargeon et al., 1994; Uller et al., 1999). In both cases, the robustness of the representations of these events is affected by repeated updates of the occluded set of objects. We conclude, based on both methodological and empirical parallels, that the representational system underlying performance on this task is shared between prelinguistic human infants and non-human primates.

The conclusions that both monkeys and human infants deploy the same representational system in the addition/subtraction violation of expectancy looking time studies, and that this system is the object indexing and tracking system of mid-level vision (Carey & Xu, 2001; Scholl & Leslie, 1999), converges with data from the spontaneous choice task (Feigenson et al., 2002a; Hauser et al., 2000). In the spontaneous choice method, rhesus monkeys and 10- to 12-month-old human infants were also given the identical task, and detailed similarities in the patterns of performance were also observed. Again, the data implicated object file representations in both the babies and the monkeys. Again, detailed similarities in the variables that affect performance on this task across the two species implicate a common representational resource with identical content, format, and computational properties.

Object files are symbols for attended individual objects. In these experiments, we can be confident that spatio-temporal information is used to individuate and trace numerical identity (sameness in the sense of *same one*), for the objects are indistinguishable in terms of properties. The monkey must establish and maintain a representation in memory of each object as it is introduced behind the screen, and then open a new object file for additional objects on the basis of spatio-temporal discontinuity with those already represented. This theoretical perspective allows us to establish exactly how object file models do and do not represent number. Unlike analog magnitude representations, there is no symbol for number. The content of the symbols in these representations is *object*. The cardinal value of the set is not explicitly represented; rather, it is represented only implicitly as there is one object file for each object. Nonetheless, these representations have numerical content. Object files track individuals and the computational processes that open and maintain object files embody criteria for individuation and numerical identity. The representational system distinguishes whether a given object is the same one or a different one from one seen earlier. Thus, models formulated over object files track whether objects are added or subtracted from attended arrays. Further, arrays may be compared on the basis of 1–1 correspondence, establishing numerical equivalence. The two-box choice studies suggest that object file representations also support ordinal computations.

We do not doubt that both non-human animals (including rhesus monkeys) and preverbal human infants have the capacity to form analog magnitude representations of number, both with and without training (Brannon & Terrace, 1998; Hauser et al., in review; Xu & Spelke, 2000). The crucial question raised by the work presented here, together with other studies, is why analog magnitude representations are not recruited in the violation of expectancy addition/subtraction studies with small sets of objects or in the two-box choice tasks. There is no reason to *expect* a principled answer to this question. Number representations may be optional in some circumstances—we do not always count the individuals in our experience. Apparently, with small numbers of individual objects, it is the objects themselves that are attended and traced through time, and in many circumstances properties of those individuals other than their number are more salient (Clearfield & Mix, 1999; Feigenson et al., 2002a; Feigenson et al., 2002b)

Even though there may not be a definitive answer to the question of when analog magnitude representations of number are recruited by an animal, there may be something more to say about the differences between the circumstances in which they are and those in which they are not. In the choice experiments, as in the present violation of expectancy looking time experiments, objects are introduced sequentially, one at a time, behind barriers or into closed containers. Thus, any mechanism that computes number representations would have to be iterative, like Meck and Church's accumulator model tested in the present experiment. Recent experiments with adults, however, suggest that analog representations of arrays of objects are often created in parallel rather than iteratively. Specifically, subjects take as long to create a numerical representation of 100 dots as they do to create a numerical representation of 20 dots (Barth, 2001). Such results favor the Church and Boradbent (1990) model for the creation of analog magnitude number representations over the Meck and Church iterative model. If this conclusion is correct, then the present experiments do not provide optimal input for the analog magnitude system of number representation. This possibility could easily be tested by giving monkeys and babies addition/subtraction violation of expectancy tasks in which larger sets are placed behind the screen in groups (e.g., $4 + 4 = 8$ or 4). If the participants succeeded under these conditions, further experiments could explore the Weber fraction signature of analog magnitude representations, sensitivity to number of updates, and so forth. Recent studies by Flombaum (2002) and Flombaum and Hauser (in prep) provide evidence that is consistent with the Church and Broadbent model, revealing that rhesus monkeys can discriminate 4 vs. 8, but not 4 vs. 6, using the expectancy violation method with appropriate controls for volume.

While the results from the present study largely converge with those from the two-box choice task in supporting object file representations, there is also a glaring difference in the two tasks. The choice task requires multiple updates—objects are placed into the container one at a time. Why are there different upper limits on performance in the two tasks? Monkeys succeed robustly at a $1+1+1$ vs. $1+1+1+1$ comparison (3 updates vs. 4 updates) while they fail in the looking time study at a $2+1+1=4$ or 3 comparison. Similarly, 10-month-old infants succeed robustly in a $1+1$ vs. $1+1+1$ comparison (2 updates vs. 3 updates) while failing in a $1+1+1=2$ or 3 comparison. The patterns of success and failure are the same across both *species*, strengthening the conclusion that the same systems of representation are at work here, but we are not certain how to think about the discrepancy across the two *paradigms*. There are at least five differences between these paradigms that may account for the lack of complete convergence. First, the looking time paradigm may be more difficult because it requires matching between two different representational formats, the imagined array formulated over object files opened sequentially and the revealed array presented in a specific spatial arrangement. The choice procedure, in contrast, allows the participant to compare two representations created in the same format—imagined arrays created by opening object files sequentially. Second, the looking time procedure forces a contrast within a spatial location (all of the action occurs at the stage) while the box choice procedure requires a contrast between two different spatial locations. Spatio-temporal information may help participants create representations of two distinct sets, each subject to limits on parallel individuation (Feigenson & Halberda, 2002). Third, other studies with both rhesus monkeys and human infants suggest that there may be differences in object knowledge as revealed by looking as opposed to action procedures. Thus, for example, studies of rhesus monkeys show that individuals correctly predict the location of a falling object behind an occluder when looking is used as an assay, but not when reaching is used (Hauser, 2001; Santos & Hauser, 2002). Fourth, the objects presented for looking and choice studies with rhesus monkeys differ. In the looking studies, we use food items that subjects have never seen in order to grab their attention, but block approach; although the items are edible, the rhesus do not know this, and thus rarely approach. In contrast, the choice studies use food items that subjects know, and thus, are motivated to approach. It is possible that the level of motivation influences the level of attention, and this in turn influences performance on these tasks. Unfortunately, it is not possible to equate these for rhesus monkeys in the wild because in the looking procedures, the use of familiar food would cause approach, therefore forcing an aborted trial, while in the choice method, using unfamiliar food or non-food would fail to elicit approach. Fifth, the attentional load on the subject differs between paradigms. In the looking studies, subjects must attend over the course of three trials while in the choice studies, they must attend for only a single trial. It is possible that the level of attention in the test trial for looking (i.e., the third trial in the session) is less than in the first and only trial of the choice procedure and that for particularly difficult discriminations, this leads to failure in the looking paradigm. Future work will focus on narrowing these possibilities.

The discovery of common systems of representation across species (in this case, preverbal human infants on the one hand and adult rhesus macaques on the other) has both methodological and theoretical import. Methodologically, it is important to show that the same procedures and dependent measures can be used across species, for this makes conclusions about common representational formats less problematic.

Once common representational resources have been established, results from one species can bear on controversies resulting from data from the other species. A modest example in the present case concerns the Marks and Cohen low-level perceptual familiarity hypothesis. It would not be advisable to use the between-subjects design of these studies with human babies (variance between babies is too high, so within subjects designs are much more powerful), so it is not possible to control for familiarity biases with infants as in the rhesus monkey studies. But if one accepts the arguments presented here that studies of infants and monkeys are tapping common representational resources, then the monkey results bear on the infant controversy. If the representational systems are the same, and monkeys' looking times are not driven by low-level perceptual matches, then neither are infants.

Theoretically, the discovery of common representational resources between nonhuman primates and human infants is relevant to the project of discovering the innate building blocks of human conceptual understanding. Object file representations, with quantificational consequences, are common to monkeys and human infants, and continue to articulate adult representations of the world as well (Carey & Xu, 2001; Hauser & Carey, 1998). The present studies, along with the two-box choice studies, add to our understanding of the mid-level object file system of representation. Specifically, the adult literature has not typically been concerned with quantitative computations over these representations (but see Trick & Pylyshyn, 1994). Nor was it known that subjects could represent two sets, as long as each was within the limits on parallel individuation. It is now important to assess whether other types of individuals, such as actions and sounds, are represented by similar attentional/short term memory systems. Recent work with cotton-top tamarins (Hauser, Dehaene, Dehaene-Lambertz, & Patalano, 2002) suggests that they can discriminate 2 vs. 3 auditory events even when the format of the familiarization material (i.e., consonant–vowel syllables) differs from the test material; following habituation to 3 consonant–vowel strings, subjects generalize to 3 tones while renewing interest to 2 tones. Similarly, work by Wynn (1996) indicates that infants discriminate 2 jumps from 3 jumps. What is not known, with certainty, is whether these representations show the set-size signature that reflects limits on parallel individuation and short-term memory or the Weber-fraction signature of number analog magnitude. Recent work on tamarins shows that individuals can discriminate speech syllables where the differences are 4 vs. 8, 4 vs. 6, and 8 vs. 12, but not 8 vs. 10 or 4 vs. 5, thereby providing evidence in favor of Weber fractions (Hauser et al., in review).

The human capacity for number representations is unsurpassed in the animal kingdom. Only humans create an explicit representational system (the integer list) with the capacity to represent natural number, and only humans create mathematics. Whereas the analog magnitude system of number representation is surely one evolutionary source for human numerical understanding, it seems likely to us that object file representations are another, distinct, evolutionary source of human numerical abilities (Carey, 2001). The system of object file representations tracks whether a given individual is the same one or a different one from another seen elsewhere and embodies computations that are equivalent to adding one. Number words are first learned for cardinal values of sets in the small number range, and the induction made when children learn to count has the computation of adding one at its core (the next count label in the count list refers to a set that is one larger than the set referred to by its predecessor). We believe, therefore, that the next critical step in our understanding of both the evolution and ontogeny of number representation will be to formulate when object file or analog magnitude systems are tapped, and how each contributes to formulating the system that is unique to humans.

Acknowledgments

For facilitating our research on Cayo Santiago, we thank Drs. John Berard, F. Bercovitch and M. Kessler; the Caribbean Primate Research Center is funded by a core Grant from the NIH (P51RR00168-37). For help running the experiments and coding the data, we thank Jonathan Flombaum, David Goldenberg, Justin Junge, Heather Pearson, Laurie Santos, Geertrui Spaepen, and Rasa Zimlicki. Funding for this work was provided by the NSF (SBR-9709744). All of the work conducted on Cayo Santiago adheres to the policies for animal care and research, and was approved by the CPRC and Harvard University (Number 97-02, 26 March, 2000).

References

Baillargeon, R., Miller, K., & Constantine, J. (1994). Ten month-old infants' intuitions about addition. *Unpublished manuscript*.

Barth, H. C. (2001). *Numerical cognition in adults: Representation and manipulation of nonsymbolic quantities.* Unpublished PhD, MIT, Cambridge.

Bercovitch, F., & Berard, J. (1993). Life history costs and consequences of rapid reproductive maturation in female rhesus macaques. *Behavioral Ecology and Sociobiology, 32*, 103–110.

Biro, D., & Matsuzawa, T. (1999). Numerical ordering in a chimpanzee (*Pan troglodytes*): Planning, executing, and monitoring. *Journal of Comparative Psychology, 113*, 178–195.

Boysen, S. T., & Bernston, G. G. (1989). Numerical competence in a chimpanzee. *Journal of Comparative Psychology, 103*, 23–31.

Boysen, S. T., Berntson, G. G., Shreyer, T. A., & Quigley, K. S. (1993). Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 107*, 208–215.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*, 746–749.

Carey, S. (2001). Evolutionary and ontogenetic foundations of arithmetic. *Mind and Language, 16*, 37–55.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. *Cognition, 80*, 179–213.

Church, R. M., & Boradbent, H. A. (1990). Alternative representations of time, number, and rate. *Cognition, 37*, 55–81.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10*, 408–411.

Cohen, L., & Marks, K. (in press). How infants process addition and subtraction events. *Developmental Science*, in press.

Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence and a new research agenda. *Behavioral and Brain Sciences, 11*, 561–579.

Dehaene, S. (1997). *The number sense*. Oxford: Oxford University Press.

Feigenson, L., & Carey, S. (in review). Tracking individuals via object files: Evidence from manual search. *Cognitive Development*.

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

Feigenson, L., Carey, S., & Spelke, E. (2002b). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44*, 33–66.

Feigenson, L., & Halberda, J. (2002). *Looking at the limits on numerical ability: Infants chunk large sets into smaller sets.* Paper presented at the International Conference on Infant Studies, Toronto, Canada.

Fernandes, D. M., & Church, R. H. (1982). Discrimination of the number of sequential events by rats. *Animal Learning and Behavior, 10*, 171–176.

Flombaum, J. (2002). *The evolution of guesstimation*, Harvard University, Honors Thesis, Cambridge.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *TICS, 4*, 59–65.

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*, 182–193.

Hauser, M. D. (1998). Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Animal Behaviour, 55*, 1647–1658.

Hauser, M. D. (2000). *Wild minds: What animals really think*. New York: Henry Holt Co.

Hauser, M. D. (2001). Searching for food in the wild: A nonhuman primate's expectations about invisible displacement. *Developmental Science, 4*, 84–93.

Hauser, M. D., & Carey, S. (1998). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In D. Cummins, & C. Allen (Eds.), *The Evolution of Mind* (pp. 51–106). Oxford: Oxford University Press.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society, London* 267, pp. 829–833.

Hauser, M. D., Dehaene, S., Dehaene-Lambertz, G., & Patalano, A. (2002). Spontaneous number discrimination of multi-format auditory stimuli in cotton-top tamarins (*Saguinus oediups*). *Cognition, 86*, B23–B32.

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

Heyes, C. M., & Huber, F. (2000). *The evolution of cognition*. Cambridge: MIT Press.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Colier, & L. P. Lipsitt (Eds.), *Advances in infancy research: Vol. 5* (pp. 65–99). Norwood: Ablex.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kawai, N., & Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature, 403*, 39–40.

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 Science, 2*, 10–18.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315*, 57–59.

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

Munakata, Y., Santos, L., O'Reilly, R., Hauser, M. D., & Spelke, E. S. (2000). Visual representation in the wild: How rhesus monkeys parse objects. *Journal of Cognitive Neuroscience, 13*, 44–58.

Murofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan troglodytes*): Subitizing and analogue magnitude estimation. *Japanese Psychological Research, 39*, 140–153.

Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month old infants. *Cognition, 62*, 121–149.

Olthof, A., Iden, C. M., & Roberts, W. A. (1997). Judgement of ordinality and summation of number by squirrel monkeys. *Journal of Experimental Psychology, 23*, 325–339.

Pepperberg, I. M. (1994). Numerical competence in an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology, 108*, 36–44.

Pylyshyn, Z. W., & Storm, R. W. (1998). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Rawlins, R., & Kessler, M. (1987). *The Cayo Santiago macaques*. New York, NY: SUNY University Press.

Rendall, D., Rodman, P. S., & Edmond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 51*, 1007–1015.

Rilling, M., & McDiarmid, C. (1965). Signal detection and fixed ratio schedules. *Science, 148*, 526–527.

Roberts, W. A., Coughlin, R., & Roberts, S. (2000). Pigeons flexibly time or count on cue. *Psychological Science, 11*, 218–222.

Rumbaugh, D. M., & Washburn, D. A. (1993). Counting by chimpanzees and ordinality judgements by macaques in video-formatted tasks. In S. T. Boysen, & E. J. Capaldi (Eds.), *The development of numerical competence. Animal and human models: Vol. 87–108*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Santos, L. R., & Hauser, M. D. (2002). A nonhuman primate's understanding of solidity: Dissociations between seeing and acting. *Developmental Science, 5*, F1–F7.

Santos, L. R., Sulkowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (*Macaca mulatta*). *Cognition, 83*, 241–264.

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

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

Simon, T. J. (1999). The foundations of numerical thinking in a brain without numbers. *Trends in Cognitive Science, 3*, 363–364.

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

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition, 36*, 97–128.

Sulkowski, G. M., & Hauser, M. D. (2000). Can rhesus monkeys spontaneously subtract? *Cognition, 79*, 239–262.

Thomas, R. K., & Lorden, R. B. (1993). Numerical competence in animals: A conservative view. In S. T. Boysen, & E. J. Capaldi (Eds.), *The Development of Numerical Competence: Animal and Human Models*. Hillsdale: Lawrence Erlbaum Associates, Publ.

Trick, L., & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review, 101*, 80–102.

Uller, C. (1997). *Origins of numerical concepts: A comparative study of human infants and nonhuman primates.* Unpublished PhD, MIT, Cambridge, MA.

Uller, C. (1999). *Assessing the infant counting limit.* Paper presented at the Biennial Meeting of the Society for Research on Child Development, Albuquerque, New Mexico.

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

Uller, C., Hauser, M. D., & Carey, S. (2001). Spontaneous representation of number in cotton-top tamarins. *Journal of Comparative Psychology, 115*, 248–257.

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*, 249–280.

Washburn, D. A., & Rumbaugh, D. M. (1991). Ordinal judgements of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science, 2*, 190–193.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*, 13–137.

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.

Wynn, K. (1996). Infants' individuation and enumeration of actions. *Psychological Science, 7*, 164–169.

Wynn, K. (1998). An evolved capacity for number. In D. Cummins, & C. Allen (Eds.), *The evolution of mind* (pp. 107–126). Oxford: Oxford University Press.

Xu, F. (1999). Object individuation and object identity in infancy: The role of spatiotemporal information, object property information, and language. *Acta Psychologia, 102*, 113–136.

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