

Visual Representation in the Wild: How Rhesus Monkeys Parse Objects

Yuko Munakata¹, Laurie R. Santos², Elizabeth S. Spelke²,
Marc D. Hauser², and Randall C. O'Reilly³

Abstract

Visual object representation was studied in free-ranging rhesus monkeys. To facilitate comparison with humans, and to provide a new tool for neurophysiologists, we used a looking time procedure originally developed for studies of human infants. Monkeys' looking times were measured to displays with one or two distinct objects, separated or together, stationary or moving. Results indicate that rhesus monkeys

used featural information to parse the displays into distinct objects, and they found events in which distinct objects moved together more novel or unnatural than events in which distinct objects moved separately. These findings show both commonalities and contrasts with those obtained from human infants. We discuss their implications for the development and neural mechanisms of higher-level vision. ■

INTRODUCTION

Physiological and anatomical studies in nonhuman primates have advanced the understanding of the primate visual system and revealed detailed homologies between human and nonhuman primate vision (Tootell, Dale, Sereno, & Malach, 1996; Sereno, Dale, & Tootell, 1995; DeYoe & Van Essen, 1988; Maunsell & Newsome, 1987; Desimone, Albright, Gross, & Bruce, 1984). Fundamental questions nevertheless remain concerning the similarities and differences between the visual representations that humans and monkeys form. For example, to what degree do nonhuman primates share our human propensity to develop taxonomies of objects, treating each living object as a member of a given species and each artifact as a tool with a specific function? Answering such questions is critical to understanding the relation between visual cognition in human and nonhuman primates and to interpreting data on the physiology and anatomy of primate vision.

Here we report four experiments investigating how semi-free-ranging rhesus monkeys form representations of and inferences about visible objects presented under natural conditions. Our experiments use methods that require no training, that allow direct comparisons with studies of humans both before and after the acquisition of language, and that could be adapted to permit simultaneous behavioral and neuronal recordings. We examine rhesus monkeys' parsing of visual

displays into objects and their sensitivity to the natural motions of those objects, using a looking time procedure that was developed for studies of human infants (e.g., Fantz, 1961), adapted for studies of object perception in human infancy (e.g., Spelke, Breinlinger, Jacobson, & Phillips, 1993), and applied to both free-ranging rhesus monkeys and captive cotton-top tamarins with considerable success (Hauser, MacNeilage, & Ware, 1996; Hauser, 1998).

The Development of Object Parsing in Humans

One motivation for the current work comes from developmental studies of object perception in human infancy. Infants perceive objects by using information about the three-dimensional arrangements and motions of visible surfaces (hereafter, *spatiotemporal information*) before they use information about the colors, textures, and curvature of surfaces (hereafter, *featural information*) (Xu, Carey, & Welch, 1999; Needham & Baillargeon, 1998; Spelke et al., 1993; Kestenbaum, Termine, & Spelke, 1987; von Hofsten & Spelke, 1985). Further, the ability to use featural information to represent distinct objects emerges at the same time as the first names for objects. This correlation raises questions about the relation between visual representation and language. Does the acquisition of a natural language, or the emergence of related symbolic capacities, lead humans to represent objects in a way that is unique to our species? Alternatively, do human and nonhuman primates form homologous object representations that emerge in humans at about 1 year of life? The current studies attempt to distin-

¹ University of Denver, ² Harvard University, ³ University of Colorado

guish these possibilities by testing adult nonhuman primates' abilities to perceive objects using featural information.

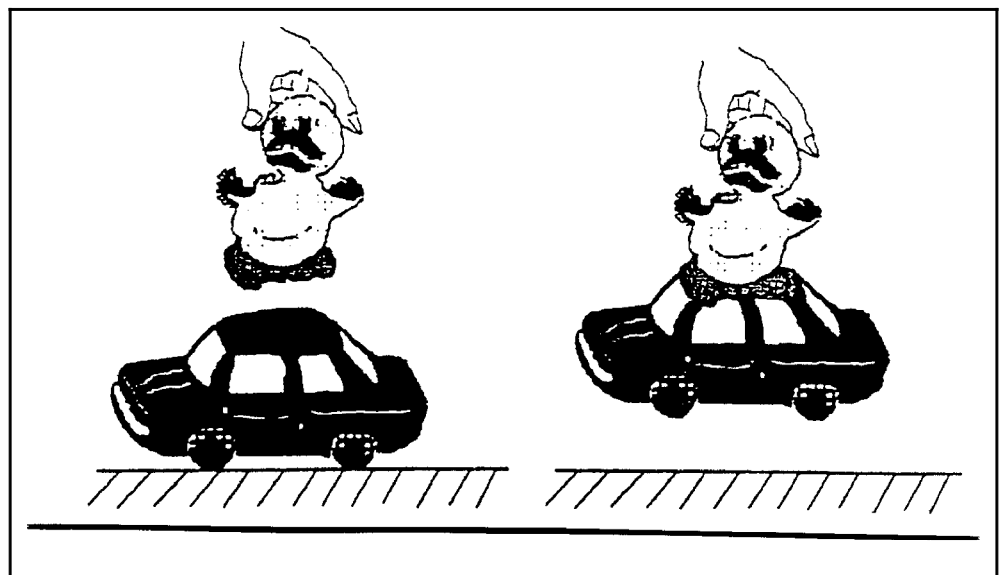
Much of the evidence concerning the development of object representations in humans comes from preferential looking experiments, in which infants are presented repeatedly with a visual display until their attention to the display declines, and then they are presented with new displays. Looking times to the new displays are measured, on the assumption that infants will look longer at displays they perceive to be novel or unnatural (Baillargeon, 1995; Fantz, 1964). These looking times therefore provide evidence concerning infants' representations of all the displays, and their (perhaps tacit) expectations about how the displays may change (Spelke, 1985).

One set of preferential looking studies provides the background for the present experiments. Xu et al. (1999) presented 10- and 12-month-old infants with an array that adults perceive as one meaningful object on top of another (e.g., a duck resting on a car) on a flat supporting surface (Figure 1). In one condition, the objects were stationary; in the other condition, the top object was moved relative to the bottom object during the initial familiarization. After infants had viewed the stationary or moving display repeatedly, they were shown two test events in which a hand grasped the top object and lifted it. In one event, the top object rose into the air while the bottom object remained on the supporting surface (left-hand side of Figure 1); in the other event, both objects moved upward together (right-hand side of Figure 1). Looking times to the outcomes of these events were compared to each other and to the looking times of infants in a baseline condition, who viewed the same outcome displays but received no initial familiarization with the objects and viewed no

object motion. Relative to this baseline measure of the displays' intrinsic attractiveness to infants, the 12-month-old infants in the main experiment looked longer at the test outcome in which the two objects had moved together, both after familiarization with moving objects and after familiarization with stationary objects. This looking pattern provides evidence that infants had parsed both of the initial displays into two bounded objects, and that they represented each object as separately movable and manipulable. The 10-month-old infants showed the same looking patterns in the condition in which the objects initially were presented in motion, but not in the condition in which the objects initially were stationary. These findings provide evidence that 10-month-old infants used spatiotemporal information, but not featural information specifying object kind, to represent object boundaries.

Xu et al.'s (1999) findings accord with research using variants of this method with younger infants and simpler object displays. Infants 3 to 5 months old have been found to parse two adjacent objects such as blocks and cones into two separately movable, manipulable units if the objects are separated in depth or undergo separate movement (Spelke, Hofsten, & Kestenbaum, 1989; Kestenbaum et al., 1987; von Hofsten & Spelke, 1985) but not if the objects are stationary, adjacent, and distinguishable only by their different surface texture, coloring, and curvature (Needham & Baillargeon, 1998; Spelke et al., 1993). Because infants are known to be sensitive to the latter featural information (see Kellman & Arterberry, 1998, for review), one interpretation of these findings is that infants' representations of objects depend on a modular system, which operates on spatiotemporal but not featural information (e.g., Scholl & Leslie, in press; Bertenthal, 1996; Spelke & Van de Walle, 1993). Alternatively,

Figure 1. Displays testing infants' sensitivity to spatiotemporal or featural information (Reprinted from *Cognition*, 70, Xu et al., Infants' ability to use object kind information for object individuation 137–166, 1999, with permission from Elsevier Science).



infants may use both spatiotemporal and featural information to parse objects, but they may be less sensitive to the latter (Needham, 1997; Johnson & Aslin, 1996). In any case, a change appears to occur at the end of the first year, when infants first reliably demonstrate the use of featural information to specify the boundaries of adjacent objects.¹ This change coincides with the emergence of the first names for objects (Xu & Carey, 1996; Xu et al., 1999).

One exception to the general rule that young infants ignore featural information in parsing objects concerns their responses to humans and human body parts, especially hands. By 6 months of age, infants treat hands as distinct from inanimate objects and have different expectations about how hands and inanimate objects should behave. In particular, infants view the motions of hands as goal-directed (Woodward, 1998), they anticipate that a hand can pick up an object but that one object cannot pick up another (Leslie, 1982), and they appreciate, on some level, that an object held by a hand requires no further support (Needham & Baillargeon, 1993). Infants' sensitivity to relationships between hands and inanimate objects may form an important precursor to the development of tool use, which is largely learned by observation in young children (Nagell, Olguin, & Tomasello, 1993; Tomasello, Kruger, & Ratner, 1993; Meltzoff, 1988). It is not known, however, whether this early-developing sensitivity is unique to humans or causally involved in later-developing object representations.

Object Representations in Monkeys

Hauser et al. (1996), Hauser (1998), and Hauser and Carey (1998) have adapted the preferential looking method for studies of object representations in both semi-free-ranging and captive monkeys. In Hauser's studies, adult monkeys are given abbreviated versions of the preferential looking experiments used with infants, with brief familiarization periods and brief, fixed-duration test trials. Monkeys typically are familiarized and tested with events involving food items, because these objects elicit high levels of spontaneous attention (Hauser et al., 1996; Hauser & Carey, 1998; though see Hauser, 1998; Hauser & Williams, submitted for cases where monkeys have been tested successfully with nonfood items). Like human infants, adult monkeys have been found to show higher levels of spontaneous looking time when they view certain events that human adults find novel or unnatural (e.g., events in which an object appears to vanish after moving behind a screen), even when care is taken to match the natural and unnatural events on a variety of other dimensions (see Hauser & Carey, 1998 for discussion). Hauser's findings suggest that the preferential looking methods developed for studies of

human infants can serve to assess object representations in monkeys as well, allowing systematic comparisons of high-level visual abilities in monkeys and humans.

Additional reasons for using preferential looking methods to test for monkeys' object representations relate to the extensive anatomical, physiological, and behavioral literature on the visual representations of nonhuman primates, and, in particular, rhesus monkeys. A wealth of studies provide evidence for homologous mechanisms subserving visual recognition in rhesus monkeys and humans (Tootell et al., 1996; Sereno et al., 1995; DeYoe & Van Essen, 1988; Maunsell & Newsome, 1987; Desimone et al., 1984). Nevertheless, progress has been slow in understanding how this recognition system works, and in gaining insights into both the commonalities and differences between human and nonhuman object representation. A standard physiological study on rhesus monkeys might demonstrate that neurons in a given brain area become active when the monkeys see a given display, after extensive training with the display. Such data, however, do not reveal the cognitive and behavioral functions of the neurons activated by the visual display, nor the ways in which monkeys use these patterns of neural activation to interpret the world. Moreover, such data make only limited contact with the rich literature in cognitive psychology and cognitive neuroscience on object representations in humans. Do monkeys parse visual displays into objects that can be separately recognized and manipulated? What expectations do monkeys form about events involving objects?

Three features of existing studies of the neural mechanisms of visual representations may have hindered progress in answering these questions. First, with the notable exception of studies of face recognition (e.g., Perrett, Mistlin, & Chitty, 1987), neurophysiological studies of visual processing in monkeys have tended to use two-dimensional, arbitrary stimuli of no ecological significance. Because monkeys' abilities to recognize objects surely have evolved to solve problems such as finding food, avoiding obstacles, and using landmarks to recognize significant places, the mechanisms subserving these abilities may not be best revealed by studies of the responses of neurons to two-dimensional geometric figures or alphanumeric characters. Second, the monkeys in neurophysiological studies typically are given extensive training with an arbitrary set of objects before physiological recording begins. It is not clear whether training regimes lasting a year or more reveal monkeys' ordinary capacities for detecting and remembering natural objects under conditions of incidental viewing, or their ad hoc strategies adopted to solve the problem at hand (see discussion in Rao, Rainer, & Miller, 1997). Third, the methods used to study monkeys often are quite

different from those used with humans. These differences complicate comparisons across species and hinder attempts to trace the evolutionary origins of human capacities in our common primate heritage (Hauser & Carey, 1998).

The present studies investigate monkeys' representations of natural, ecologically significant objects—food items—undergoing ecologically significant events—grasping and lifting. Moreover, they use a method that has been used extensively with humans of all ages, requires no training, and can be administered to free ranging as well as captive animals. For these reasons, the studies promise to shed light on the representations of objects that monkeys form spontaneously, paving the way for simultaneous behavioral and neurophysiological studies of the mechanisms of object representations in untrained animals, under conditions allowing systematic comparison with humans.

In the present studies, we used a variant of the method of Xu et al. (1999) to investigate whether untrained, free-ranging adult rhesus monkeys, with a mature visual system but no spontaneous tool use and, at best, a limited capacity for language, perceive object boundaries as older human infants do. Monkeys were presented with displays containing two novel food objects that were stationary and adjacent. They then viewed the lifting of the top object or of both objects together on two separate trials, and their looking times to the outcomes of these events were recorded. In one condition, the lifting of the objects was accomplished by a single human hand that grasped only the top object. In a second condition, the lifting was accomplished by two hands, each grasping one object and moving the objects together. If monkeys perceived the

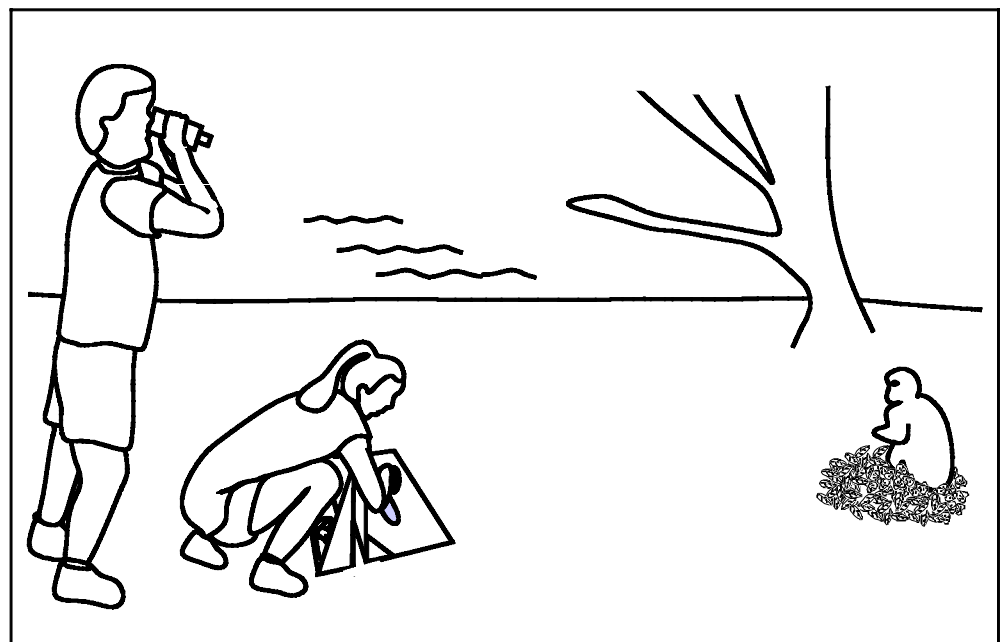
object boundaries as older infants do, they should look longer at the outcome of the event in which the two objects moved together in the one-hand events. This tendency might be attenuated in the two-hand events, because each object was lifted by a supporting hand. Such findings would suggest that human and nonhuman primates have homologous representations of objects as movable and manipulable units, and that both species distinguish hands from other objects and are sensitive to the functions of hands in supporting and moving objects.

EXPERIMENT 1

In our first experiment, monkeys were presented with two novel food items—either a pumpkin and a piece of ginger root or a pepper and a sweet potato—on the floor of a stage (Figure 2). After a monkey had observed one item sitting on top of the other item for at least 1 sec, a hand grasped the top item and lifted it. On different trials, either the top item moved alone while the bottom item remained on the stage floor (“separate trial”), or the two items moved together (“together trial”). At the end of the movement, the hand and objects remained stationary at their final positions for 10 sec, and the monkey's looking time was recorded. Looking times to the two test displays were compared to investigate whether monkeys looked longer at the event outcome on the “together” trial, a preference that would suggest that they perceived two separately movable objects in the original display.

To investigate monkeys' sensitivity to the role of human hands in supporting and manipulating objects,

Figure 2. Experimental set-up: The experimenter positioned the apparatus 2–5 m away from the test monkey. The apparatus consisted of a stage and a screen that could block the view of the stage and store the food stimulus items for the study. The screen was placed behind the stage during test trials, as shown in the figure.



the test events were presented in two different ways to different groups of subjects. For half the subjects, both events were produced by a single hand that grasped only the top food object (hold-top). In this condition, the bottom food object appeared to human observers to rest naturally on the display floor on the “separate” trial and to move unnaturally with the top object on the “together” trial. For the remaining subjects, each food object was grasped by a different hand, such that both objects appeared to be adequately supported on both trials (hold-both). Comparisons of monkeys’ looking preferences across these conditions should reveal whether monkeys take account of the support function of human hands in representing object motions.

Results

Figure 3a and b present the findings from this experiment. Monkeys looked longer at “together” events (5.1 sec, $SE = .3$ sec) than at “separate” events (4.1 sec, $SE = .3$ sec), $F(1, 57) = 9.2, p < .005$. This effect was equally strong in the hold-top and hold-both conditions, yielding no interaction of Condition by Display ($F = 0$). Of the 59 monkeys tested, 42 looked longer at “together” events and 17 looked longer at “separate”, $\chi^2 = 11, p <$

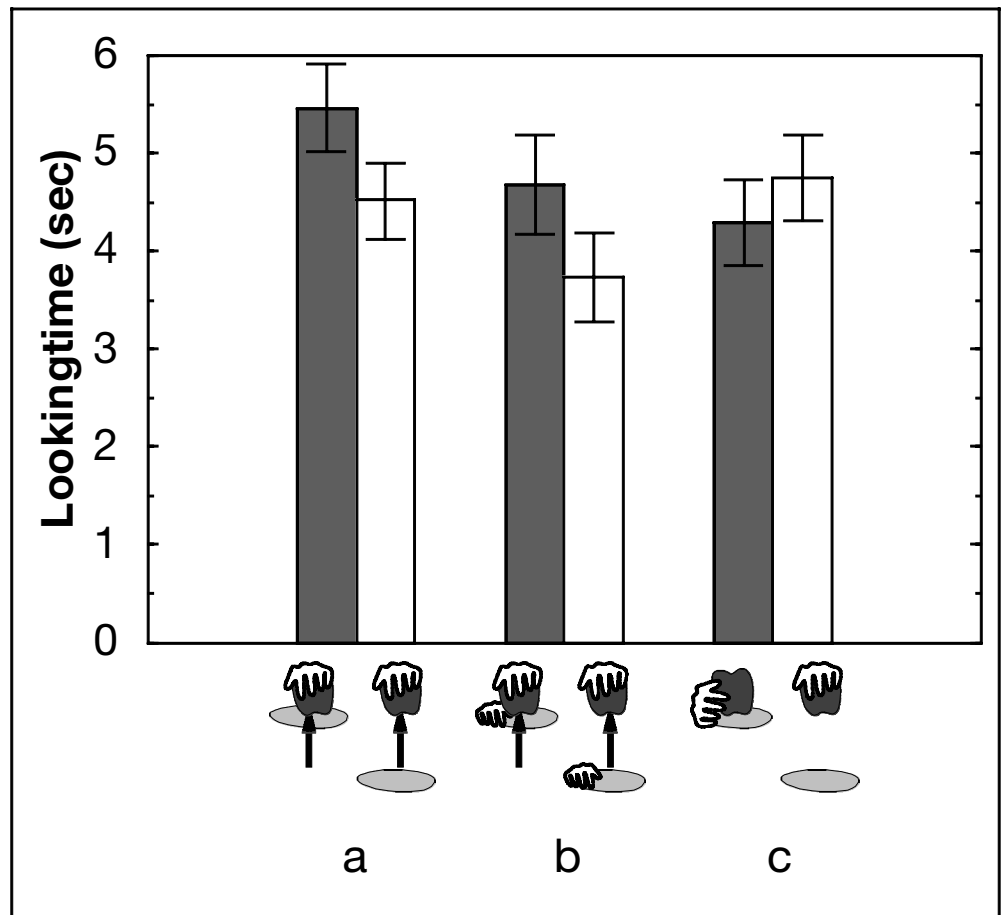
.005. The main effect of Condition was not statistically significant ($F < 2$).

Discussion

These results show that rhesus monkeys look longer when two bounded objects move together than when one of the two objects moves separately from the other. Like human infants, monkeys appear to parse arrays into bounded objects, and they represent these objects as independently movable and manipulable. Moreover, monkeys and infants alike appear to look longer at events in which two perceptually bounded objects move and behave as a single unit, suggesting that they find such events to be novel or unnatural. Monkey and human object representations therefore appear to be similar and to be testable by similar methods, in accord with previous findings (Hauser et al., 1996).

The experiment also revealed two differences between the object representations of adult monkeys and young human infants. First, infants as young as 6 months take account of the supporting role of hands in lifting and moving objects, but the monkeys in Experiment 1 showed no sensitivity to hands. Informal observations suggested that the monkeys were highly attentive to the food objects but oblivious to the human

Figure 3. Results from Experiments 1 and 2. Rhesus monkeys look longer when two distinct objects move together than when one of the two objects moves separately, both (a) when one hand holds and lifts the top object (Experiment 1, hold-top condition) and (b) when two hands hold both objects (Experiment 1, hold-both condition). In contrast, (c) rhesus do not distinguish “together” from “separate” displays in their looking times when the distinct objects are stationary (Experiment 2).



hands that held and manipulated them. Second, infants below 12 months do not consistently perceive the boundary between two adjacent objects that are stationary, even when the objects belong to different familiar kinds. Because the initial displays in the present studies contained objects that were adjacent and underwent no relative motion, the present findings suggest that the rhesus monkeys used featural information to parse the visual display into distinct objects.

According to this object-parsing interpretation, two factors are critical to the rhesus' responses: the distinct features of the objects and their distinct or common motions. Because these two factors were not manipulated separately in the test displays in Experiment 1, however, there are two alternative accounts of these findings, each of which discredits one of the factors critical to the object-parsing account. According to one alternative account, monkeys simply find displays with distinct features in spatial proximity more interesting than displays with distinct features in more distant locations. Monkeys may look longer at the outcome of the "together" trial because there are more features clustered together than in the outcome of the "separate" trial; the motion of the objects may be completely irrelevant to monkeys' looking behavior. Experiment 2 tests this alternative interpretation of the findings by comparing monkeys' looking times to the same two outcome displays with no preceding motion. If the alternative interpretation is correct, then monkeys should show a preference for the "together" outcome over the "separate" outcome in Experiment 2, as in Experiment 1. If the object-parsing interpretation is correct, in contrast, monkeys should not show the same preference for the "together" outcome in Experiment 2.

The second alternative interpretation of the data from Experiment 1 is that monkeys' looking times to different event outcomes depend on how much motion preceded those outcomes. According to this account, monkeys looked longer at the outcome of the "together" trial because a greater volume of food moved during the event that preceded the recording of their looking time. On this interpretation, the distinct features of the objects and the representation of object boundaries are irrelevant to monkeys' looking behavior. Experiment 3 tests this alternative interpretation by presenting monkeys with "together" and "separate" trials in which a single object moves as a whole or splits apart. If the alternative interpretation is correct, then monkeys should show a preference for the "together" outcome over the "separate" outcome in Experiment 3, as in Experiment 1, because in both experiments the "together" outcome was preceded by motion of a greater volume of food. In contrast, the object-parsing interpretation predicts that monkeys will not show the same preference for the "together" outcome in Experiment 3 without the distinct featural information.

EXPERIMENT 2

A new group of monkeys was presented with the two outcome displays from Experiment 1, without any prior exposure to the food objects or to their motions. Because hands were found not to influence monkeys' looking patterns in Experiment 1, all the outcome displays presented two food items held by one hand. On one trial (together), a hand held both food items in the air by grasping both objects at once, one atop the other. On the other trial (separate), a hand held one food item in the air while the other food item rested on the display floor. Looking times to the two test outcomes were compared to each other and to monkeys' looking times to the same outcome displays in Experiment 1. If the looking preference for the "together" outcome in Experiment 1 reflected monkeys' parsing of the initial arrays into two objects and their expectation that the two objects would move independently, then that preference should be absent or attenuated in Experiment 2.

Results

Figure 3c presents the principle findings of Experiment 2. With stationary objects, monkeys looked equally at "together" events (4.3 sec, $SE = .4$ sec) and "separate" events (4.7 sec, $SE = .4$ sec), $F(1, 27) = 1.5, p > .22$. Of the 28 monkeys tested, nine looked longer at the "together" event, 18 looked longer at the "separate" event, and one individual looked equally at both events ($\chi^2 = 3.0$, nonsignificant). The "tie" data point from the single individual was dropped for the χ^2 analysis.

An analysis comparing looking times in Experiments 1 and 2 revealed a significant interaction between trial type (together vs. separate) and experiment (1 vs. 2), $F(1, 85) = 7.4, p = .008$. Monkeys showed a greater looking preference for the "together" outcome display in Experiment 1.

Discussion

In Experiment 2, rhesus monkeys looked no longer at a display in which two different objects were held in the air together than at a display in which one object was held in the air while the other object rested on the display floor. These findings contrast with the results from Experiment 1, which focused on monkeys' looking times to these same displays after prior exposure to two adjacent objects and to their common or separate motions. These findings challenge one alternative explanation for the looking preferences in Experiment 1 and support our object-parsing interpretation of those looking preferences, whereby monkeys parse visual displays into distinct objects based on featural information and find events in which distinct objects move together more interesting than events in which distinct objects move separately.

However, the second alternative interpretation, whereby monkeys look longer at displays presenting the outcomes of events in which a greater volume of food has moved, could account for the data from both experiments. Experiment 3 tests this alternative with food displays that are parsed, by human infants and adults, into a single object that either moves as a whole or breaks apart. If monkeys' looking times to event outcomes depend on the volume of food in motion that preceded each outcome, then they should look longer at an outcome in which a whole food object has moved than at an outcome in which half of the object has moved. In contrast, if monkey's looking times depend on their parsing of visual displays into bounded objects, then they should show different looking preferences at the outcomes of events that involve one versus two objects.

EXPERIMENT 3

As in Experiment 1, monkeys were presented with a display of food sitting on a stage floor, a hand grasped the top of the food display and lifted either just the top half of the food or all of the food into the air, and then the display remained stationary while looking times to the event outcomes were recorded. In con-

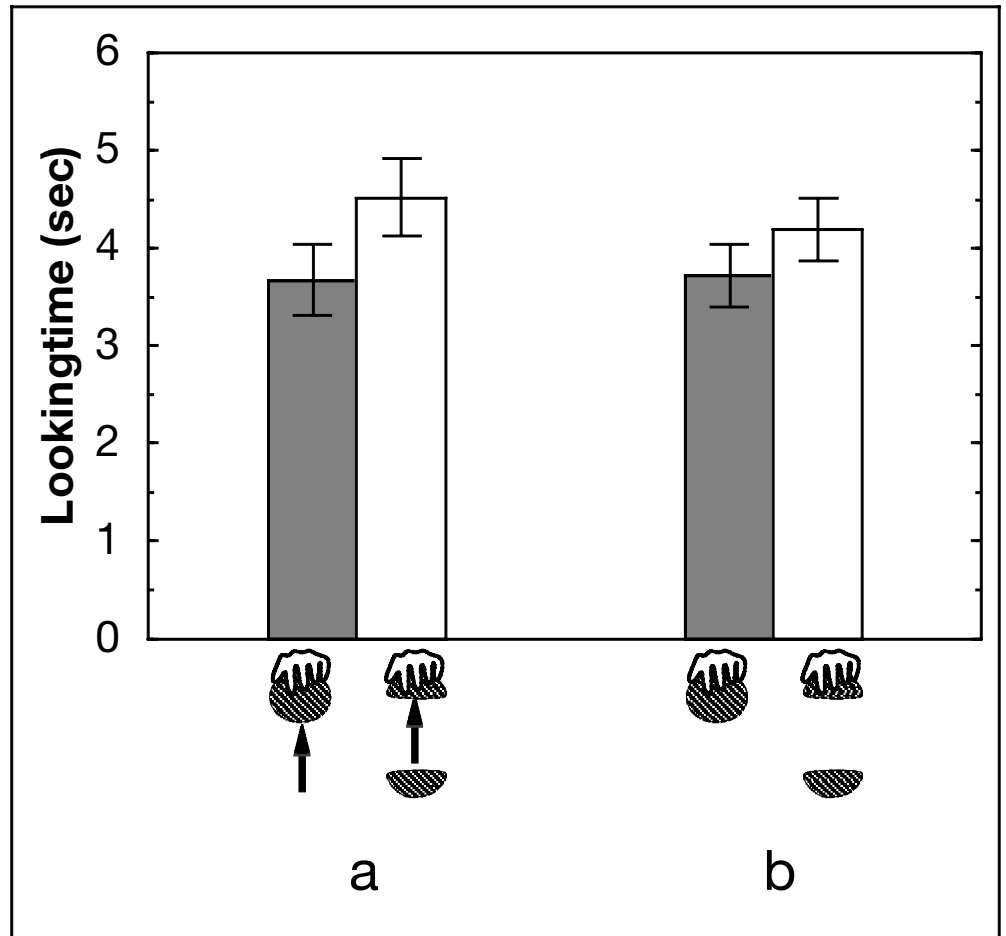
trast to Experiment 1, however, each display contained a single food item—a lemon or an orange pepper—that either broke into two pieces (separate) or moved as a whole (together). Looking times at the event outcomes were compared to one another and to the looking times of the monkeys in Experiment 1, to investigate whether monkeys' preferences between the event outcomes depends on the volume of food that is lifted or on the monkeys' parsing of the food into distinct objects.

Results

Figure 4a presents the principal findings of Experiment 3. Monkeys showed a nonsignificant trend toward looking longer at the "separate" outcome display (4.5 sec, $SE = .4$ sec) than at the "together" outcome display (3.7 sec, $SE = .4$ sec), $F(1, 29) = 3.2, p = .08$. Of the 30 monkeys tested, 19 looked longer at the "separate" display and 11 looked longer at the "together" display ($\chi^2 = 2.1$, nonsignificant).

In contrast, an analysis comparing looking times in Experiments 1 and 3 revealed a significant interaction between test display (together vs. separate) and experiment (1 vs. 3), $F(1, 100) = 8.5, p = .004$. Monkeys

Figure 4. Results from Experiments 3 and 4. Rhesus look nonsignificantly longer when one object appears in two pieces than when it appears together as a whole, both (a) when the object is moved (Experiment 3) and (b) when it is stationary (Experiment 4).



looked longer at the outcome of the “together” event in Experiment 1 than in Experiment 3.

Discussion

When monkeys were presented with events in which either a single food item moved as a whole or half the object moved independently of the rest, they did not look longer at the event outcome that followed motion of a greater food volume. Indeed, monkeys showed a marginally significant tendency in the opposite direction, looking longer at the outcome of the event in which the object broke apart. Looking preferences between the “together” and “separate” trials differed significantly from the preferences shown in Experiment 1, in which the events involved two distinct objects. These findings accord with the thesis that monkeys use featural information to parse visual scenes into objects, represent each object as separately movable and manipulable, and look longer at events in which two distinct objects move together.

Nevertheless, one of the alternative accounts could be revised to account for this collection of data. Perhaps monkeys have a preference both for event outcomes that follow the motion of more food, and for event outcomes that reveal the inside of a food object. According to this revised account, monkeys in Experiment 1 looked longer following an event in which two distinct objects moved together because of their preference for more stuff moving. This preference was not evident in Experiment 3, because it competed with an intrinsic preference for the outcome display from the “separate” trial. Because the inside of the lemon or pepper was visible following the “separate” event of Experiment 3 but not following either the “together” event in Experiment 3 or either event in Experiment 1, a preference for viewing the inside of a food object would produce a greater preference for the “separate” outcome display in Experiment 3 than in Experiment 1.

Experiment 4 tests this revised account by presenting a new group of monkeys with the outcome displays of the “together” and “separate” events from Experiment 3, with no prior presentation of any objects or motion. According to the revised account, monkeys should show a stronger preference for the “separate” event in Experiment 4 than in Experiment 3, because only Experiment 3 would invoke the competing preference for more stuff moving in the “together” event. According to the original, object-parsing account, the preference for the “separate” event in Experiment 4 will not exceed that in Experiment 3. If the monkeys in Experiment 3 expect single objects to move as cohesive units, then preference for the outcome of the “separate” event might be greater in Experiment 3 than in Experiment 4. If monkeys have no expectations about the cohesive or non-cohesive motion of food objects, then preferences should be the same in the two experiments.

EXPERIMENT 4

Experiment 4 used the outcome displays of Experiment 3 and the method of Experiment 2: Monkeys were presented with one stationary display in which a hand held a whole food object in the air (together), and one stationary display in which a hand held the top half of the food object in the air while the bottom half of the food object rested on the display floor (separate). Looking times to the two displays were compared to each other and to the looking times of the monkeys in Experiment 3, who viewed the same displays following presentation of the whole object and two different patterns of motion.

Results

Figure 4b presents the principal findings of Experiment 4. Monkeys looked equally at “together” events (3.7 sec, $SE = .3$ sec) and “separate” events (4.2 sec, $SE = .3$ sec), $F(1, 42) = 1.5$, $p = .2$. Of the 43 monkeys tested, 16 looked longer at the “together” event and 27 looked longer at the “separate” event ($\chi^2 = 2.8$, nonsignificant).

The analysis comparing looking times in Experiments 3 and 4 revealed a significant main effect of trial type: monkeys looked longer at the “separate” outcome display (4.3 sec, $SE = .3$ sec) than at the “together” outcome display (3.7 sec, $SE = .2$ sec), $F(1, 71) = 4.5$, $p < .05$. Of the 73 monkeys tested in Experiments 3 and 4, 46 looked longer at the “separate” outcome display and 27 looked longer at the “together” outcome display, $\chi^2 = 4.9$, $p < .05$.

Discussion

In Experiment 4, rhesus monkeys showed a nonsignificantly smaller preference for the “separate” display, in which a single food item appeared in two pieces, than their counterparts in Experiment 3. This finding provides evidence against the thesis that monkeys’ looking times depend on a preference for the outcomes of events involving the motion of more food stuff, combined with an intrinsic preference for the separated outcome display with one object. They instead support the object-parsing interpretation of the results from Experiment 1. Monkeys appear to use featural information to parse visual displays into distinct objects, and they find events in which distinct objects move together more novel or less natural than events in which distinct objects move separately.

The findings of Experiments 3 and 4 provide no clear evidence concerning monkeys’ expectation that single food items will move cohesively. If monkeys had such an expectation, then the subjects in Experiment 3 should have looked longer at the “separate” display than those in Experiment 4, because the “separate” display in Experiment 3 followed an event in which a single object

broke apart and moved noncohesively. Although the data from Experiments 3 and 4 tend in this direction, no reliable differences were obtained between the looking preferences in the two experiments. Reliable preferences for the outcomes of noncohesive motions have been observed both with human infants and with human adults tested with similar methods and with displays of simple artifacts (Spelke et al., 1989, 1993; Kestenbaum et al., 1987). The absence of a clear effect of cohesiveness in Experiments 3 and 4 may reflect either a species difference or a difference in object domain: Artifacts are more apt to move cohesively than is food, which breaks apart both due to decay, cutting, or eating. Such conclusions cannot be drawn from the present experiments, however, because of the equivocal findings.

GENERAL DISCUSSION

Four experiments provide evidence that rhesus monkeys spontaneously parse arrays of adjacent food items into distinct objects, and that they represent these objects as separately movable and manipulable. Monkeys looked longer at the outcomes of events in which two previously stationary, adjacent objects moved as one unit than at the outcomes of events in which one of the objects moved separately from the other. This preference was not attributable to any intrinsic preference for the former event outcome or to any preference for an outcome that followed a greater amount of motion. Instead, it provides evidence that the monkeys represented the common motion of the two distinct objects as more novel or surprising than the independent motion of those objects.

The present findings suggest broad similarities between the object representations formed by human and nonhuman primates, and between the ways in which those representations are used to support inferences about objects' movability. The well known, detailed homologies between the lower-level visual mechanisms of human and nonhuman primates (Tootell et al., 1996; Sereno et al., 1995; DeYoe & Van Essen, 1988; Maunsell & Newsome, 1987; Desimone et al., 1984) therefore appear to extend to higher-level mechanisms for parsing objects and interpreting object motions. In addition, our findings provide evidence that adult monkeys and human infants show similar behavioral responses to object motions, with heightened visual exploration of motions that are novel or surprising. These findings complement previous results showing that rhesus monkeys, cotton-top tamarins, and human infants show similar looking preferences for events in which objects are occluded or behave in anomalous ways (e.g., Hauser et al., 1996; Hauser, 1998).

Differences in Sensitivity to Hands

Our studies also reveal two differences between the object representations formed by adult rhesus monkeys and young human infants. First, human infants take

account of the actions of human hands in analyzing the motions and support relations among objects. When human infants see an inanimate object rise into the air in a display that includes a human hand, they show a novelty reaction if the hand and object are spatially separated but not if the hand is grasping the object (Needham & Baillargeon, 1993; Leslie, 1984). Monkeys, in contrast, showed no sensitivity to the supporting role of hands in Experiment 1. Their novelty reaction to the common rising motion of two objects was equally strong when no hand contacted the bottom object (an event that implies that the two objects were connected) and when hands contacted each of the objects (an event that implies no connection between the objects).

We see two plausible accounts of the observed differences in sensitivity to hands. First, human infants' greater sensitivity to the supporting role of hands may reflect a species difference in the use of hands, specifically in the manipulation of inanimate objects. Because human infants and human adults manipulate objects more than other primates do, human infants may have more opportunities to learn about hand-object support relations than do other species. A second possibility, not mutually exclusive from the first, is that humans are innately predisposed to attend to the ways in which inanimate objects are manipulated by other humans, which in turn contributes to both infants' abilities to learn rapidly about tools and ultimately to humans' superior tool use.²

Differences in the Use of Object Features for Boundaries

The second difference between the object representations of adult monkeys and young human infants concerns the use of object features such as surface coloring and shape as information for object boundaries. Adult monkeys and human infants above 11 months of age use featural information to perceive object boundaries; in contrast, infants below 11 months of age do not reliably exhibit this ability. Various factors have been proposed to underlie the developmental change observed in humans. Some factors focus on perceptual development, with behavioral changes attributed to infants' emerging abilities to use image features such as edge alignment and texture similarity to group portions of the visual field into units directly (e.g., Kellman & Arterberry, 1998; Needham, 1998). In contrast, other factors focus on the development of higher level processes, with behavioral changes attributed to an emerging ability to represent objects as members of kinds, and an emerging propensity to use object features such as surface coloring and shape as information for the kinds to which specific objects belong (Needham & Modi, 2000; Xu & Carey, 1996). Further, this change may be driven by the acquisition of verbal labels for the objects (Xu & Carey, 1996).

Corresponding to these two interpretations of the developmental change in humans are two different interpretations of monkeys' performance in the present studies: Monkeys may have perceived the object boundaries by categorizing each object as a different kind of food, or they may have perceived the boundaries grouping together elements in the visual scene in accord with their colors, textures, and alignment relationships.

There is compelling data suggesting that monkeys represent the category of food, such that they are likely to have "food kind" representations. First, monkeys in the present studies were strongly attentive to food items and occasionally attempted to approach and take them: behaviors often observed with familiar foods and rarely observed with familiar nonfood objects. This was true even though they had no prior experience with these particular food items. Second, experiments by Santos, Hauser, and Spelke (in preparation) suggest that monkeys given evidence that a novel object is food (by observing a person eating part of it) subsequently approach that object, an odorless replica of that object, and other objects of the same color and texture as the original object but of a different shape. In contrast, monkeys do not approach these objects when they are given evidence that the initial object is not food (by observing a person putting the object in her ear rather than her mouth). This finding suggests that monkeys categorize novel objects as kinds of food in terms of properties such as their colors and textures. If perceptible properties of the present stimulus objects allowed monkeys to perceive correctly that these objects were food, then monkeys' propensity to categorize objects as the same foods only when they share a common color and texture would lead them to perceive each display of two (differently colored and textured) foods as containing two distinct objects.

Whatever the reason for monkeys' successful use of featural information to perceive object boundaries, the existence of this capacity in rhesus monkeys casts doubt on the thesis that this ability either depends on, or gives rise to, any uniquely human ability to represent objects. Humans do represent objects in unique ways, for we have unparalleled abilities to build and use complex tools and to communicate about objects with unique symbols for thousands of object kinds. The sources of our uniqueness, however, do not clearly appear in the contexts that have been used thus far to assess object representations in human infants.

Steps Toward a Cognitive Neuroscience of Natural Object Representation

Although our experiments focus strictly on behavioral measures and functional analyses, we believe their greatest potential lies in the contributions they can make to understanding the neural basis of object representation. Rhesus monkeys are one of the most intensively studied

species in the neuroanatomy and neurophysiology of vision, and such studies have provided evidence for extensive homologies between their visual systems and those of humans. Our experiments contribute to this literature in three ways. First, they suggest that rhesus monkeys and humans have similar higher visual mechanisms for representing objects and interpreting object motions. The origin of these similarities remains an open question, with likely contributions from both genetically encoded homologies in the underlying neural architectures and similar experiential histories interacting with similar neural learning mechanisms.

Second, our experiments provide evidence that the object representations of monkeys and humans can be assessed by nearly identical tasks. Moreover, these tasks require no training and so allow assessment of the representations that humans and monkeys develop and use spontaneously, rather than less naturalistic representations that may have been developed specifically for solving experimental tasks over months of training on those tasks (see discussion in Rao et al., 1997). Finally, these tasks can be applied not only to adult animals but to infants. Indeed, the preferential looking method was developed for use with infant humans and monkeys (Fantz, 1961), and it has been used to study lower-level visual functions in both species (see Kellman & Banks, 1998 for review). The method therefore should be ideal for investigating the neural architecture subserving visual cognition in both species.

Third, our experiments offer a behavioral task that can readily be adapted for simultaneous behavioral and neural recordings in monkeys. Preferential looking methods have been used successfully both with semi-free-ranging rhesus monkeys and with captive cotton-top tamarins (e.g., Hauser et al., 1996; Hauser, 1998). In preliminary research, they have yielded similar findings with rhesus monkeys tested with stabilized heads and implanted electrodes (Munakata, Miller, & Spelke, unpublished). In the future, therefore, cognitive neuroscientists should be able to use these methods to probe the neural mechanisms of object representations in untrained monkeys whose experience with objects can be precisely controlled, and to compare the functional properties of those mechanisms directly to those of human infants with varying degrees of experience. Such studies should prove a valuable complement to studies of the neural mechanisms of object representations in adult humans using the combined approaches of cognitive psychology and functional brain imaging.

More specifically, the studies reported in this paper could serve as the starting point for physiological studies probing the cognitive and behavioral functions of neurons activated by visual displays. As a first step, one could ask whether the extensively studied object coding neurons in the inferotemporal cortex (Tanaka, 1996; Perrett et al., 1987; Baylis, Rolls, & Leonard, 1985) are

responsible for the behavioral results found in our experiments. The finding that monkeys encode two-object displays as two separate objects leads to the prediction that inferotemporal neurons will respond similarly to each object in a one-object display and in a two-object display, with a possible reduction in response to the latter display due to competition from the different object representations. It is possible, however, that monkeys distinguish the two objects in earlier stages of processing, parsing the display based on contiguous regions of the same general color and texture, without this parsing being clearly reflected in object-level representations. These alternatives could be distinguished by recording a population of inferotemporal responses to one of our two-object displays and to each of the two objects separately. If the two-object responses were different from the sum or average of the responses to the two separate objects, this would suggest that monkeys encode the two-object displays in a different manner than the separate objects at the level of the inferotemporal cortex.

If this first experiment showed that inferotemporal neurons encode two-object displays in terms of the two separate objects, one could next manipulate factors that influence object perception and measure the neural correlates. For example, spatiotemporal cues such as common motion may make monkeys more likely to perceive a two-object display as a single object (e.g., Kellman & Spelke, 1983), and elimination of color differences may make them less likely to do so (Santos, Hauser, & Spelke, in preparation). One could thus measure both the behavioral (looking time) and electrophysiological consequences of such manipulations, and compare the results to those from experiments without this preexposure.

Another ideal candidate for converging exploration focuses on the representations underlying abilities to perceive combinations of objects in terms of the separable components: Do such abilities stem from representations of distinct perceptual features or from representations of distinct object kinds? The nature of object representations is a matter of considerable debate in the electrophysiological and related literature (e.g., Sugihara, Edelman, & Tanaka, 1998; Logothetis & Sheinberg, 1996; Tanaka, 1996; Tarr & Bulthoff, 1995; Biederman & Cooper, 1992; Biederman & Gerhardstein, 1995), and issues related to the kinds/features distinction have been discussed, in a somewhat different terminology. For example, Logothetis and Sheinberg (1996) posit that different levels of categorization could be used to organize object representations, from more specific visual feature-based representations to more abstract-kind representations. Electrophysiological recordings have demonstrated that a given visual object is represented in different ways along a rough hierarchy of processing pathways, from more specific, low-order featural representations

to more abstract, invariant, categorical representations (e.g., Desimone & Ungerleider, 1989). Objects such as those used in our displays could be recognized as distinct based on features at lower levels or categories at higher levels. Alternatively, even the lowest level of object representations may be organized into different kind categories, as suggested by the existence of face-specific representations in both rhesus monkey and human visual areas (Kanwisher, McDermott, & Chun, 1997; Perrett et al., 1987). Though most of the explanations for face-specific representations focus on the unique perceptual properties of faces, rather than a more general categorical organization of the object recognition system, a categorical organization is still possible (Caramazza, 1998). Thus, objects could be categorized as different kinds at the earliest levels of featural processing.

One could further explore these issues in physiological studies by presenting monkeys with different visual forms of a single food category (e.g., bananas that are sliced, mashed, peeled, unpeeled, green, brown, and yellow), and visually similar forms from different food categories (e.g., a green banana and a cucumber). If inferotemporal representations encode information at the level of kinds, the first condition should elicit similar responses in the inferotemporal neurons, whereas the second should elicit different responses. In contrast, if inferotemporal representations encode information at the level of features, the first condition should elicit different responses, whereas the second should elicit similar responses.

In such ways, our understanding of object representation, and in turn, of humans' unique tool and symbol use, may be enhanced by converging efforts at the behavioral and physiological levels of analysis. The methods reported in this paper—used extensively with humans of all ages, requiring no training, and applicable to free-ranging as well as captive animals—could play an instrumental role in this process.

METHODS

Experiment 1

Participants

Subjects were 59 semi-free-ranging rhesus monkeys living on the island of Cayo Santiago, Puerto Rico. Approximately half the subjects were adult males (age >4 years) and half adult female (age >3 years). Subjects were tested opportunistically whenever they were encountered in a setting with few other monkeys or distractions (e.g., not involved in, or near to, a fight), and when they remained in a seated position long enough for us to present our stimuli. Monkeys occasionally changed positions between trials. In these cases, testing resumed if and when monkeys relocated to another seated position within a couple of minutes. An

additional 21 monkeys were tested but did not provide data for the analyses, due to either position changes that did not allow testing to resume (20 monkeys) or experimenter error (1 monkey).

Apparatus and Displays

The experimental apparatus consisted of a stage and a screen constructed from white foam core (Figure 2). The 60 × 30-cm floor and 60 × 40-cm back of the stage were attached at a right angle by triangular supports (12-cm height × 7-cm base) attached to the sides of the stage. The 60 × 40-cm screen had a 60 × 15-cm base supporting small aluminum pans containing the food stimuli for the study. The base and pans were attached to the screen at a right angle by large triangular supports (40-cm height × 15-cm base) that occluded both the base and the food.

The objects were four foods of contrasting shapes, colors, and textures, with sizes that made them easily graspable by a single human hand: a green pepper (7-cm tall × 8-cm diameter), a brown sweet potato (7-cm tall × 7.5-cm wide by 17-cm long), a miniature orange pumpkin (7-cm tall × 8-cm diameter), and a segment of tan ginger root (4.5-cm tall × 12-cm wide × 15-cm long). None of these items grew on the island or were brought there either as provisions for the monkeys or as food for the research team; all items therefore were unfamiliar to the subjects. In one display, the green pepper rested on top of the sweet potato. In the other display, the pumpkin rested on top of the ginger root.

Design

Each monkey was presented with one “together” trial and one “separate” trial, each involving a different pair of food items. Twenty-eight monkeys were tested in a hold-top condition in which the experimenter held only the top object with one hand during each event, and 31 monkeys were tested in a hold-both condition in which the experimenter held both objects with both hands. Within each of these conditions, the pairing of objects (pepper/potato vs. pumpkin/ginger) and trial types (together vs. separate) and the order of test trials were orthogonally counterbalanced across monkeys.

Procedure

All testing was conducted by one experimenter and one camcorder operator; a test began when the investigators located a monkey who was seated in a quiet spot. The experimenter positioned the apparatus 2–5 m away from the test monkey with the screen in front of and blocking the monkey’s view of the stage, and the camcorder operator began to videotape the monkey from behind the display (Figure 2). The experimenter then raised the

screen to reveal an empty stage and immediately lowered it. Each test trial then proceeded as follows: The experimenter raised the screen to reveal one food item sitting atop a second food item. The experimenter checked that the monkey had fixated the objects, and then she lifted the top object approximately 30 cm in 1 sec. In “together” events, the bottom object moved with the top object; in “separate” events, the bottom object remained on the floor of the display. In the hold-top condition, the experimenter held only the top object from above with the right hand (the two objects were attached with toothpicks invisible to the monkeys). In the hold-both condition, the experimenter held the top object from above with the right hand and the bottom object from the side and bottom with the left hand. After lifting the object(s), the experimenter called “Count” and the camcorder operator began counting 10 sec on the camcorder display. The experimenter held the object(s) stationary until the camcorder operator called “Done” to signal the end of the 10-sec trial. The experimenter then lowered the screen. This procedure has been successfully used in previous looking time experiments on this population (e.g., Hauser et al., 1996).

Each monkey received one “together” and one “separate” trial. These two trials were separated by two additional trials unrelated to the present studies and involving the stationary presentation of other food items (carrots and squash). For most monkeys, trials were separated by an intertrial interval of 3–5 sec and the entire experiment lasted a couple of minutes. For monkeys who repositioned themselves between trials, the intertrial intervals were longer but never exceeded a couple of minutes.

Coding and Analysis

Two coders blind to the hypotheses and conditions of the experiment viewed the videotaped trials frame-by-frame to determine how long monkeys observed each of the event outcomes. On each trial, coding began just after the objects came to rest, as signaled by the experimenter’s voice on the videotape, and ended 10 sec later. Four of the monkeys were coded by both coders; the correlation between their judgments of total looking time on each trial was .93. Looking times were analyzed by a 2 × 2 ANOVA with Condition (hold-top vs. hold-both) as a between-subjects factor and Display (together vs. separate) as the within-subjects factor.

Experiment 2

Participants

Subjects were 28 monkeys from the same population as in Experiment 1. An additional 10 monkeys were tested

but did not provide data for the analyses, due to either position changes that did not allow testing to resume (9 monkeys) or experimental error (1 monkey).

Apparatus and Stimuli

The apparatus was similar to that in Experiment 1, except that the stage was somewhat smaller (back = 45 × 30 cm, floor = 45 × 30 cm) and the screen was slightly taller (base = 45 × 15 cm; face = 45 × 45 cm). The food objects and object positions were the same as in the outcome displays for Experiment 1. On the “separate” trial, the position of the experimenter’s hand was the same as in the “separate” trial of the hold-top condition of Experiment 1. On the “together” trial, the experimenter’s right hand grasped the two objects simultaneously from the side and supported them in the same positions as on the “together” trials for both conditions of Experiment 1.

Design

The design was the same as in Experiment 1, except that all subjects were run in a single-hand condition.

Procedure

Each trial began when the experimenter lifted the screen to reveal the objects currently held in the air. As the screen was raised, the experimenter called “Count” and the camcorder operator began counting 10 sec on the camcorder display. In all other respects, the procedure was the same as in Experiment 1.

Coding and Analysis

A single coder blind to the conditions of the experiment scored the videotapes. Trials for 10 subjects were coded by a second observer and the correlation between judgments of both observers was .98. As in previous studies (Hauser et al., 1996; Hauser, 1998), videos were acquired onto a computer using Adobe Premiere software and a Radius Videovision board. Coding began and ended as for Experiment 1.

Looking times in Experiment 2 were analyzed by a one-way ANOVA with Display (together vs. separate) as the within-subjects factor. A further ANOVA with the additional factor of Experiment compared the looking patterns of the monkeys in Experiment 2 to those in Experiment 1.

Experiment 3

Participants

Subjects were 30 monkeys from the same population as in Experiments 1 and 2. An additional 13 monkeys were

tested but did not provide data for the analyses, due to either position changes that did not allow testing to resume (12 monkeys) or experimental error (1 monkey).

Apparatus and Displays

The apparatus was identical to that of Experiment 2. The displays were the same as in the hold-top condition of Experiment 1, except for the objects: a yellow lemon and an orange pepper, oriented vertically. On “together” trials, a whole object appeared on the display floor, oriented vertically, and a hand grasped its top half and lifted the object into the air. On “separate” trials, two halves of an object with a horizontal cut through the middle appeared on the display floor in the same orientation, and a hand grasped the top half and lifted it into the air while the bottom half remained on the display floor. At the start of the “separate” trial, the cut in the object was detectable by adults but inconspicuous. At the end of the trial, small portions of the inside of the object were visible from the monkey’s station-point.

Design, Procedure, Coding, and Analyses

The design and procedure were the same as in Experiment 1, except that only one condition (hold-top) was administered and only one object (together or separate) was displayed. The coding and analyses were the same as in Experiment 2.

Experiment 4

Participants

Subjects were 43 monkeys from the same population as in Experiments 1–3. An additional 27 monkeys were tested but did not provide data for the analyses, due to either position changes that did not allow testing to resume (26 monkeys) or experimental error (1 monkey).

Apparatus and Displays

These were the same as in Experiment 3, except that the food object never appeared on the display floor and was not grasped and lifted.

Design, Procedure, Coding, and Analyses

These were the same as in Experiment 2.

Acknowledgments

The research was supported by McDonnell-Pew Postdoctoral Fellowships to Yuko Munakata and Randall O’Reilly; an NSF predoctoral fellowship and Harvard University McMaster’s funds to Laurie Santos; NIH grant R37-HD23103 to Elizabeth Spelke; and an NSF Young Investigator Award to Marc

Hauser. We thank members of the Cognitive Development Center at the University of Denver for feedback on drafts of this article. We thank Elliott Blass, Amy Jackendoff, Katie Liu, Cory Miller, Marianne Moon, Bridget Spelke, and Fei Xu for assistance with conducting and coding the studies. We thank the CPRC (PHS grant P51RR00168-38), and Drs. M. Kessler, F. Bercovitch, and J. Berard for helping secure the Cayo Santiago facilities.

Reprint requests should be sent to Yuko Munakata at the Department of Psychology, University of Denver, 2155 S. Race Street, Denver, CO 80208, USA, or via e-mail to munakata@du.edu.

Notes

1. In some cases, infants younger than 1 year have demonstrated abilities to use featural information to parse simple, adjacent objects into separable units (Needham & Baillargeon, 1997; Needham, 1998). Further work is needed to determine why such abilities are not reliably observed in infants during the first year of life.

2. A third account for the differential sensitivity to hands may be quickly rejected. Human infants' greater response to hands is not due to any greater sensitivity of the preferential looking task with human infants, relative to adult rhesus monkeys, because we found monkeys to be more sensitive than human infants to featural information for objects in the present experiments. A fourth interpretation is possible though not very likely. Human infants' greater sensitivity to hands in our experiments may depend on the use of human hands—perhaps monkeys would show similar sensitivity to hands if tested with monkey hands. However, rhesus monkeys show similar physiological responses to human hands and monkey hands (Rizzolatti, Fadiga, Fogassi, & Gallese, 1999; di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992), suggesting that rhesus monkeys' observed insensitivity to the supporting role of hands in our studies is unlikely to be an artifact of the use of human hands.

REFERENCES

- Baillargeon, R. (1995). Physical reasoning in infancy. In M. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge: MIT Press.
- Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex of the superior temporal sulcus of the macaque monkey. *Brain Research*, *342*, 91–102.
- Bertenthal, B. I. (1996). Origins and early development of perception, action, and representation. *Annual Review of Psychology*, *47*, 431–459.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 121–133.
- Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bulthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1506–1514.
- Caramazza, A. (1998). Domain-specific knowledge systems in the brain: The animate–inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34.
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.
- Desimone, R., Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology*, vol. 2 (pp. 267–299). New York: Elsevier.
- DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*, 219–226.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*, 176–180.
- Fantz, R. (1961). The origin of form perception. *Scientific American*, *204*, 66–72.
- Fantz, R. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.
- Hauser, M. D. (1998). A nonhuman primate's expectations about object motion and destination: The importance of self-propelled movement and animacy. *Developmental Science*, *1*, 31–37.
- 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*. Oxford: Oxford University Press.
- Hauser, M. D., MacNeilage, P., & Ware, M. (1996). Numerical representations in primates. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 1514.
- Hauser, M. D., & Williams, T. (submitted). A nonhuman primates' expectations about invisible displacement: Two procedures, two different systems of knowledge?
- Johnson, S. P., & Aslin, R. N. (1996). Perceptions of object unity in young infants: The rules of motion, depth, and orientation. *Cognitive Development*, *11*, 161–180.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302.
- Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge: MIT Press.
- Kellman, P. J., & Banks, M. S. (1998). Infant visual perception. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language*, 5th ed (pp. 103–146). New York: Wiley.
- Kellman, P. J., & Spelke, E. (1983). Perception of partially occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.
- Kestenbaum, R., Termine, N., & Spelke, E. S. (1987). Perception of objects and object boundaries by 3-month-old infants. *British Journal of Developmental Psychology*, *5*, 367–383.
- Leslie, A. M. (1982). The perception of causality in infants. *Perception*, *11*, 173–186.
- Leslie, A. M. (1984). Infant perception of a manual pick-up event. *British Journal of Developmental Psychology*, *2*, 19–32.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.
- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, *10*, 363–401.
- Meltzoff, A. S. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development*, *59*, 217–225.
- Nagell, K., Olguin, R. S., & Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology*, *107*, 174–186.
- Needham, A. (1997). Factors affecting infants' use of featural information in object segregation. *Current Directions in Psychological Science*, *6*, 26–33.
- Needham, A. (1998). Infants' use of featural information in the segregation of stationary objects. *Infant Behavior and Development*, *21*, 47–76.

- Needham, A., & Baillargeon, R. (1993). Intuitions about support in 4.5-month-old infants. *Cognition*, *47*, 121–148.
- Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, *62*, 121–149.
- Needham, A., & Baillargeon, R. (1998). Effects of prior experience on 4.5-month-old infants' object segregation. *Infant Behavior and Development*, *21*, 1–23.
- Needham, A., & Modi, M. (2000). Infants' use of prior experiences with objects in object segregation: Implications for object recognition in infancy. In H. Reese (Ed.), *Advances in child development and behavior*, vol. 27 (pp. 99–133).
- Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neurosciences*, *10*, 358–364.
- Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, *276*, 821.
- Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives Italiennes de Biologie*, *137*, 85.
- Scholl, B., & Leslie, A. (in press). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylyshyn (Eds.), *What is cognitive science?* Oxford: Blackwell.
- Sereno, M. I., Dale, A. M., & Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889.
- Spelke, E. (1985). Preferential looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 323–363). Norwood, NJ: Ablex.
- Spelke, E., Breinlinger, K., Jacobson, K., & Phillips, A. (1993). Gestalt relations and object perception: A developmental study. *Perception*, *22*, 1483–1501.
- Spelke, E., & Van de Walle, G. A. (1993). Perceiving and reasoning about objects: Insights from infants. In N. Eilan & R. A. McCarthy (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 132–161). Oxford: Blackwell.
- Spelke, E. S., Hofsten, C. V., & Kestenbaum, R. (1989). Object perception in infancy: Interaction of spatial and kinetic information for object boundaries. *Developmental Psychology*, *25*, 185–196.
- Sugihara, T., Edelman, S., & Tanaka, K. (1998). Representation of objective similarity among three-dimensional shapes in the monkey. *Biological Cybernetics*, *78*, 1.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.
- Tarr, M. J., & Bulthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1494.
- Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*, 495–552.
- Tootell, R. B. H., Dale, A. M., Sereno, M. I., & Malach, R. (1996). New images from human visual cortex. *Trends in Neurosciences*, *19*, 481–489.
- von Hofsten, C., & Spelke, E. (1985). Object perception and object-directed reaching in infancy. *Journal of Experimental Psychology: General*, *11*, 198–212.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.
- Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*, 111–153.
- Xu, F., Carey, S., & Welch, J. (1999). Infants' ability to use object kind information for object individuation. *Cognition*, *70*, 137–166.