From Fragments to Geometric Shape:

Changes in Visual Object Recognition Between 18 and 24 Months

Linda B. Smith
Indiana University

Abstract

Visual object recognition is foundational to processes of categorization, tool use, and real-world problem solving. Despite considerable effort across many disciplines and many specific advances, there is no comprehensive or well-accepted account of this ability. Moreover, none of the extant approaches consider how human object recognition develops. New evidence indicates a period of rapid change in toddlers' visual object recognition between 18 and 24 months that is related to the learning of object names and to goal-directed action. Children appear to shift from recognition based on piecemeal fragments to recognition based on geometric representations of three-dimensional shape. These findings may lead to a more unified understanding of the processes that make human object recognition as impressive as it is.

Keywords

visual object recognition; object name learning; development; visuomotor development

Human visual object recognition is fast, robust, and successful in the service of a variety of different tasks. For example, people routinely recognize the dog whose nose is sticking out from the blanket; they recognize deck chairs and kitchen chairs as chairs; and they recognize their favorite cup as their own. The range of these abilities suggests that visual object recognition depends not on a single process but on several distinct processes (Peissig & Tarr, 2007). This article is primarily concerned with the visual representations that support recognition at the level of basic categories—for example, the processes that enable us to recognize easy chairs, lawn chairs, and rocking chairs as chairs. New developmental evidence indicates a significant shift in the nature of these representations in children between the ages of 18 and 24 months.

Two classes of theories—so-called object-based and view-based theories of adult object recognition—are relevant to the developmental findings. The best-known theory on the object-based side, Biederman's (1987) recognition-by-components (RBC) account proposes that humans form internal representations that are geometric models of objects' shapes and that are internally manipulated via processes analogous to mental rotation (Marr & Nishihara, 1978). These representations, built from a primitive set of geometric volumes

(Fig. 1a) capture the whole object's geometric structure independent of one's viewing perspective. The alternative class of theories explains object recognition not in terms of the geometry of three-dimensional shapes but rather in terms of picture-like (and therefore view-dependent) images (see Peissig & Tarr, 2007, for review). One of these, Ullman's (2007) "fragment" account (Fig. 1b), specifically explains recognition at the basic-category level in terms of class-specific fragments. In this account, horses, for example, are recognized via piecemeal and category-specific local fragments such as the the ears, legs, and head shape.

Both theories have been widely tested in studies of object recognition in adults and each captures important phenomena. Neither approach has seriously considered the developmental origins of visual object recognition, however. The findings reviewed below suggest an early shift from more fragment-based object recognition to recognition based on geometric shape.

CHANGE BETWEEN 18 AND 24 MONTHS

The development of visual object recognition is relatively unstudied, and there are many open questions (see Kellman, 2001, for a review). The representations and processes that underlie the visual recognition of three-dimensional real-world objects have been particularly neglected. There are, however, reasons to expect that these processes will undergo significant change during development: First, many models of high-level vision recognition, as well as behavioral and neuroscience studies, indicate a formative role for category learning (see Peissig & Tarr, 2007). Second, evidence from a different domain of visual recognition, that of face recognition, indicates a protracted course of development and learning that stretches from infancy into adolescence (e.g., Maurer, Le Grand, & Mondloch, 2002). Object recognition might well show a similarly long developmental trajectory.

Growth in Geometric Representations

The period between 18 and 24 months is an interesting one with respect to the development of object recognition, because this is when children acquire a substantial number of object names, names that refer to categories of things that are principally alike in their in shape (Gershkoff-Stowe & Smith, 2004). Moreover, there is a well-documented increased attention to object shape over material properties such as color and texture during this same period (Gershkoff-Stowe & Smith, 2004). Motivated by these findings, I (Smith, 2003) wanted to know if 18- to 24-month-old children could recognize common objects given minimal information about their geometric shape, the same kind of information posited by Biederman's RBC model to account for adult recognition. Smith compared children's recognition of two kinds of holdable and manipulable three-dimensional objects: geometric "caricatures" (Fig. 2a) constructed from two to four volumes arranged to represent overall shape but without any fine-grained detail, color, or textural information; and richly detailed typical examples (Fig. 2b). There were two measures of object recognition. In the nonlinguistic play task, children were presented with caricatures or detailed examples and their play actions were scored as indicating recognition. For example, pretending to brush hair with a brush, eat the toy slice of pizza, or take a picture with the camera were scored as indicating recognition, whereas banging, stacking, or rolling were not. In the name-

comprehension task, children were shown three objects and asked to indicate one (e.g., "show me the camera").

Both tasks yielded the same result: Older children recognized the shape caricatures as well as they did the detailed instances. Younger children did not. They recognized only the detailed examples but not the caricatures. These results provide two new insights: First, representations of global geometric shape—of the kind posited in some theories of adult object recognition—are sufficient, in and of themselves, for object recognition in 2-year-olds, just as they are in adults. The fact that the older children in this sample—who are, after all, very young—recognized the caricatures just as well as they did the detailed examples shows that these children have abstracted the geometric structure of the shapes of common objects. Second, the additional fact that younger children recognized the detailed examples in both the nonlinguistic and linguistic tasks but failed to recognize the shape caricatures suggests a change in the representations that support visual object recognition. In particular, geometric representations of object shape appear to first emerge between 18 and 24 months, a result that has been replicated in additional studies (Jones & Smith, 2005; Pereira & Smith, 2009; Son, Smith, & Goldstone, 2008).

Why might this developmental period be crucial for developing whole-object representations of geometric shape? One possibility is that object-name learning itself plays a role. I (Smith, 2003; see also Pereira & Smith, 2009) specifically examined the relation between children's recognition of the shape caricatures and the number of object names in children's vocabularies and found that known object names were a better predictor of children's shape-caricature recognition than was age. Jones and Smith (2005) provided further evidence by showing delays in visual object recognition in children with delayed vocabulary development.

The codevelopment of object-name learning and visual-object recognition does not unambiguously indicate that learning object names promotes the development of geometric representations of shape. Indeed, there is evidence for the opposite dependency; that is, the emergence of more abstract representations of object shape may facilitate learning basic-level categories. In an artificial-name-learning study, Son, Smith, and Goldstone (2008) showed that teaching object names with minimalist geometric representations led 18-month-olds to make more category-appropriate extensions (to richly detailed new examples) than did training with richly detailed examples. They concluded that more abstract representations of geometric shape enable more category-appropriate (and mature) lexical generalizations.

Fragments First?

The youngest children in these studies recognized the richly detailed examples as well as did the more advanced children, but apparently did not do so via whole-object geometrical representations. Do younger children, then, recognize objects via their parts or features? A programmatic series of studies by Rakison and colleagues (see Rakison, 2003, for review) suggest they might. These studies show that 14- and 22-month-old children base category decisions on highly salient parts (such as legs and wheels) and not on overall shape. Pereira and I (Pereira & Smith, 2009) provided a direct test of the idea that early object recognition

is fragment based. Using a forced-choice task, we compared 18- to 24-month-old children's ability to recognize objects given local featural details with their ability to recognize geometric caricatures. The study compared four kinds of objects (all three-dimensional, holdable things), also shown in Figure 2: geometric caricatures with localized category-specific features (Fig. 2c), geometric caricatures with no added features (Fig. 2d), scrambled caricatures with localized category-specific features (Fig. 2e), and scrambled caricatures with no features (Fig. 2f). Younger children (and those with few object names in their productive vocabulary) recognized the objects whenever category-specific features were present, regardless of the appropriateness of the overall shape. The older children, in contrast, performed well whenever overall geometric structure was appropriate to the named category. In brief, early object recognition appears to be based on local and category-specific features but, with development, to become more dependent on geometric shape.

A ROLE FOR ACTION?

The shift from picture-like fragments to three-dimensional shape may depend on learning object names, as a name provides a mechanism through which multiple views and examples may be integrated. However, action provides an alternative and independent pathway for building unified whole-object representations. One recent study that supports a role for action examined the relation between the development of visual completion and manual exploration in infants. Given a view of just one side of a never-before-seen object, adults have strong expectations about the geometric structure of the whole (Tse, 1999). For example, when shown the view in Figure 3a, adults expect a rotation of that object to reveal a solid volume (3b) and not a shell (3c). Visual completion implies unified representations of three-dimensional objects. Soska, Adolph, and Johnson (in press) recently showed that these expectations emerge in infants between 5 and 8 months and are related to individual infants' opportunities to manually explore objects, opportunities that increase during that time period as infants develop sufficient postural control to sit and manually play with objects for extended periods of time.

Manual exploration develops into goal-directed actions—banging and stacking objects and inserting them into openings—that require the alignment and coordination of multiple objects. These actions both depend on and may direct attention to geometric structure. In one relevant study, Örnkloo and von Hofsten (2007) examined toddlers' ability to insert objects of particular shapes into shape-matching holes, an experimental variant of everyday shape-sorter toys. Children 18 months and younger rarely oriented an object properly for insertion and rarely succeeded. In contrast, children 22 months and older were much more successful in orienting an object with respect to its hole and in inserting it. Moreover, these older children typically made appropriate adjustments of hand shape and orientation—the adjustments necessary to grasp and rotate an object for insertion—prior to picking up the to-be-inserted object. This indicates that they were able to plan their actions based on the relevant geometric properties of objects in relation to their holes. This ability to represent geometric structure in planning actions emerges in the same developmental period as the emergence of geometric representations in visual object recognition, a potentially meaningful hint of a developmental connection.

A final result that suggests a possible role for action in the development of visual object recognition concerns attention to an object's major axis of elongation (the axis of maximal length). The principle axis is an object-centered property that provides a *viewpoint independent* means for aligning objects and their mental representations (for example, for aligning two different views of the same three-dimensional shape). An object's axis of elongation in relation to the body is also important for grasping, for goal-directed actions, and for predicting an object's likely path of motion (Sekuler & Swimmer, 2000). Consistent with these observations, Smith (2005) showed that experience in moving objects along constrained paths (but not the experience of merely watching objects move along those paths) altered 2-year-olds' perception (and/or memory) of object shape. In these experiments, the children were given a three-dimensional ball-like object. With one hand, the children then moved the object repeatedly along either a vertical path or a horizontal path. Children who moved the object vertically subsequently judged its shape to be more vertically extended than it really was, and children who moved it horizontally judged it to be more horizontally extended. The direction of action apparently highlighted the corresponding visual axis (perhaps by highlighting the vertical or horizontal direction as the visual frame of reference, see Sekuler & Swimmer, 2000) thus altered the object's perceived shape. Manually moving objects is a physical analogue to mental manipulations of whole-object representations and could, therefore, be critical to the development of object-based representations.

FUTURE DIRECTIONS

Visual object recognition is a fundamental skill—an important component of category learning, problem solving, and goal-directed action. Because it is so fundamental to so many aspects of human behavior, and because it must be robust under many different viewing conditions, it seems likely that humans employ multiple, partially redundant processes. Two such processes that are evident in adults are recognition via local and fragmented features and recognition via minimal geometric structure. Developmental studies suggest that both of these kinds of object recognition are evident in very young children but that recognition via fragments develops early and that representation and recognition of objects in terms of whole-object geometric shape emerges later, specifically between 18 and 24 months.

The period between 18 and 24 months is one of considerable change in patterns of connectivity in the brain (Stiles, 2008) and is also a period of remarkable behavioral change. The developmental findings reviewed here suggest links between the emergence of whole-object representations of shape, object-name learning, and goal-directed action. Future work needs to focus on the precise nature of these relationships as well as their links to brain development. Understanding these relations may be particularly beneficial to understanding several developmental disorders, including autism-spectrum disorders and specific language impairment, as disruptions in the development of object recognition have been implicated (Behrman, Thomas, & Humphreys, 2006; Jones & Smith, 2005) in both cases, and may be a contributing factor to delays in language learning.

The possible role of action in the development of whole-object representations is intriguing. Contemporary research in cognitive neuroscience indicates a coupling between brain regions

involved in visually recognizing objects and those involved in producing actions (e.g., Chao & Martin, 2000). That is, *visual* presentations of objects with which people typically have had extensive motor interactions appear to automatically activate the cortical motor areas responsible for those actions. There are several open questions about these links, including whether they play a role in the development of visual recognition. One possibility is that such links are neural correlates of mere co-occurrence, so that although the motor regions are activated in response to visual stimuli, perhaps as preparation for action, they play no direct role in the visual recognition of the objects which instead may be based purely on visual information unrelated to action. Nonetheless, because action structures the visual input, the visual information generated by actions on objects may be critical to developing object representations. A second possibility is that these motor activations themselves feed back on and directly influence development in visual regions.

In conclusion, a complete theory of human visual object recognition will be a developmental theory. Understanding changes in human object recognition between 18 and 24 months— and how they relate to object-name learning and to action on objects—appears essential to achieving such a complete theory.

Acknowledgments—

REFERENCES

Behrman M, Thomas C, & Humphreys K (2006). Seeing it differently: Visual processing in autism. Topics in Cognitive Science, 10, 258–264.

Biederman I (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115–117. [PubMed: 3575582]

Chao L, & Martin A (2000). Representation of manipulable man-made objects in the dorsal stream. Neuroimage, 12, 478–484. [PubMed: 10988041]

Gershkoff-Stowe L, & Smith LB (2004). Shape and the first hundred nouns. Child Development, 75, 1098–1114. [PubMed: 15260867]

Jones SS, & Smith LB (2005). Object name learning and object perception: A deficit in late talkers. Journal of Child Language, 32, 223–240. [PubMed: 15779885]

Kellman PJ (2001). Separating processes in object perception. Journal of Experimental Child Psychology, 78, 84–97. [PubMed: 11161426]

Marr D, & Nishihara HK (1978). Representation and recognition of spatial organization of three–dimensional shapes. Proceedings of the Royal Society of London, Series B, 200, 269–294.

Maurer D, Le Grand R, & Mondloch CJ (2002). The many faces of configural processing. Trends in Cognitive Sciences, 6, 255–260. [PubMed: 12039607]

Örnkloo H, & von Hofsten C (2007). Fitting objects into holes: On the development of spatial cognition skills. Developmental Psychology, 43, 404–416. [PubMed: 17352547]

Peissig JJ, & Tarr MJ (2007). Visual object recognition: Do we know more now than we did 20 years ago? Annual Review of Psychology, 58, 75–96.

Pereira A, & Smith LB (2009). Developmental changes in visual object recognition between 18 and 24 months of age. Developmental Science, 12, 67–80. [PubMed: 19120414]

Rakison DH (2003). Parts, motion, and the development of the animate-inanimate distinction in infancy In Rakison DH & Oakes LM (Eds.), Early category and concept development: Making sense of the blooming, buzzing confusion (pp. 159–192). New York: Oxford University Press.

Sekuler AB, & Swimmer MB (2000). Interactions between symmetry and elongation in determining reference frames for object perception. Canadian Journal of Experimental Psychology, 54, 42–56. [PubMed: 10721238]

Smith LB (2003). Learning to recognize objects. Psychological Science, 14, 244–250. [PubMed: 12741748]

Smith LB (2005). Action alters shape categories. Cognitive Science, 29(4), 665–679. [PubMed: 21702789]

Son JY, Smith LB, & Goldstone RL (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. Cognition, 108, 626–638. [PubMed: 18565504]

Soska KC, Adolph KE, & Johnson SP (in press). Systems in development: Motor skill acquisition facilitates 3D object completion. Developmental Psychology.

Stiles J (2008). Fundamentals of brain development: Integrating nature and nurture. Cambridge, MA: Harvard University Press.

Tse PU (1999). Volume completion. Cognitive Psychology, 39, 37–68. [PubMed: 10433787]

Ullman S (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends in Cognitive Sciences, 11, 58–64. [PubMed: 17188555]

Recommended Reading

Bloom P (2000). How children learn the meanings of words. Cambridge, MA: MIT Press.A comprehensive and engaging book on early word learning, including the learning of object names.

Palmeri TJ, & Gauthier I (2004). Visual object understanding. Nature Reviews Neuroscience, 5, 1–13.A comprehensive review of the cognitive neuroscience literature on object recognition.

Rakison DH, & Woodward AL (2008). New perspectives on the effects of action on perceptual and cognitive development. Developmental Psychology, 44, 1209–1213. [PubMed: 18793054] An overview on the growing evidence that action plays a driving role generally in cognitive development.

Smith LB (2005). Shape: A developmental product In Carlson L & VanderZee E (Eds.), Functional features in language and space (pp. 235–255). Oxford, England: Oxford University Press.A review of the literature on the relation between object-name learning and attention to shape.
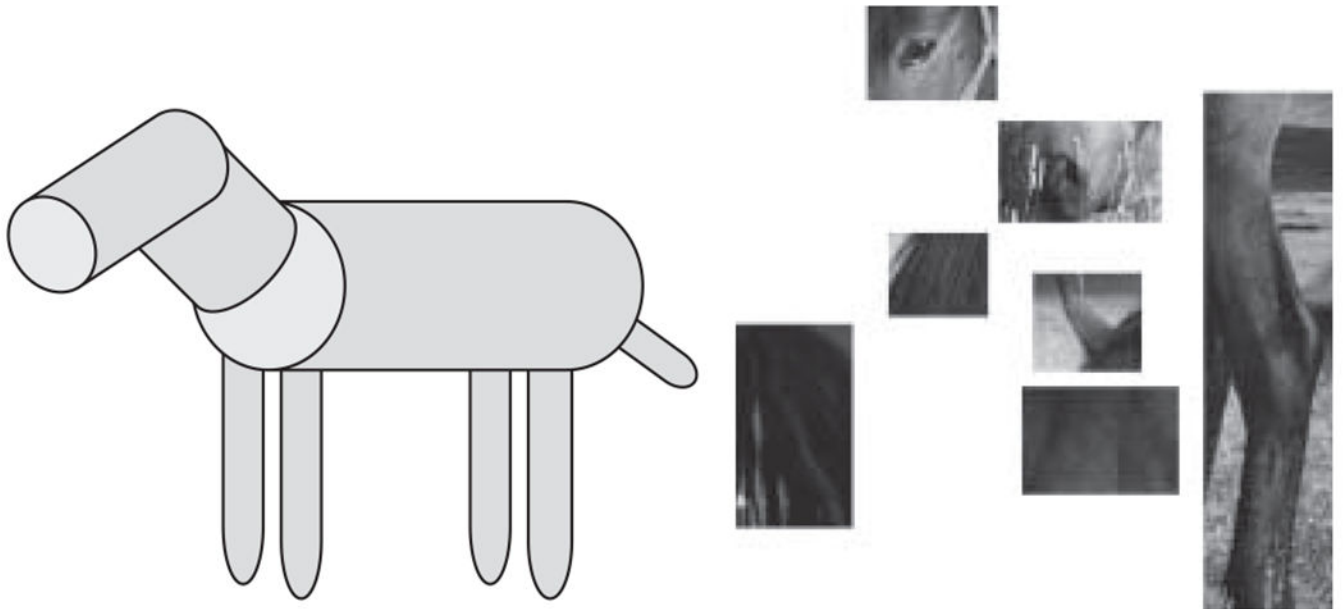
Fig. 1.
Two representations of a horse: (a) a sparse geometric representation and (b) a category-specific-fragment representation.
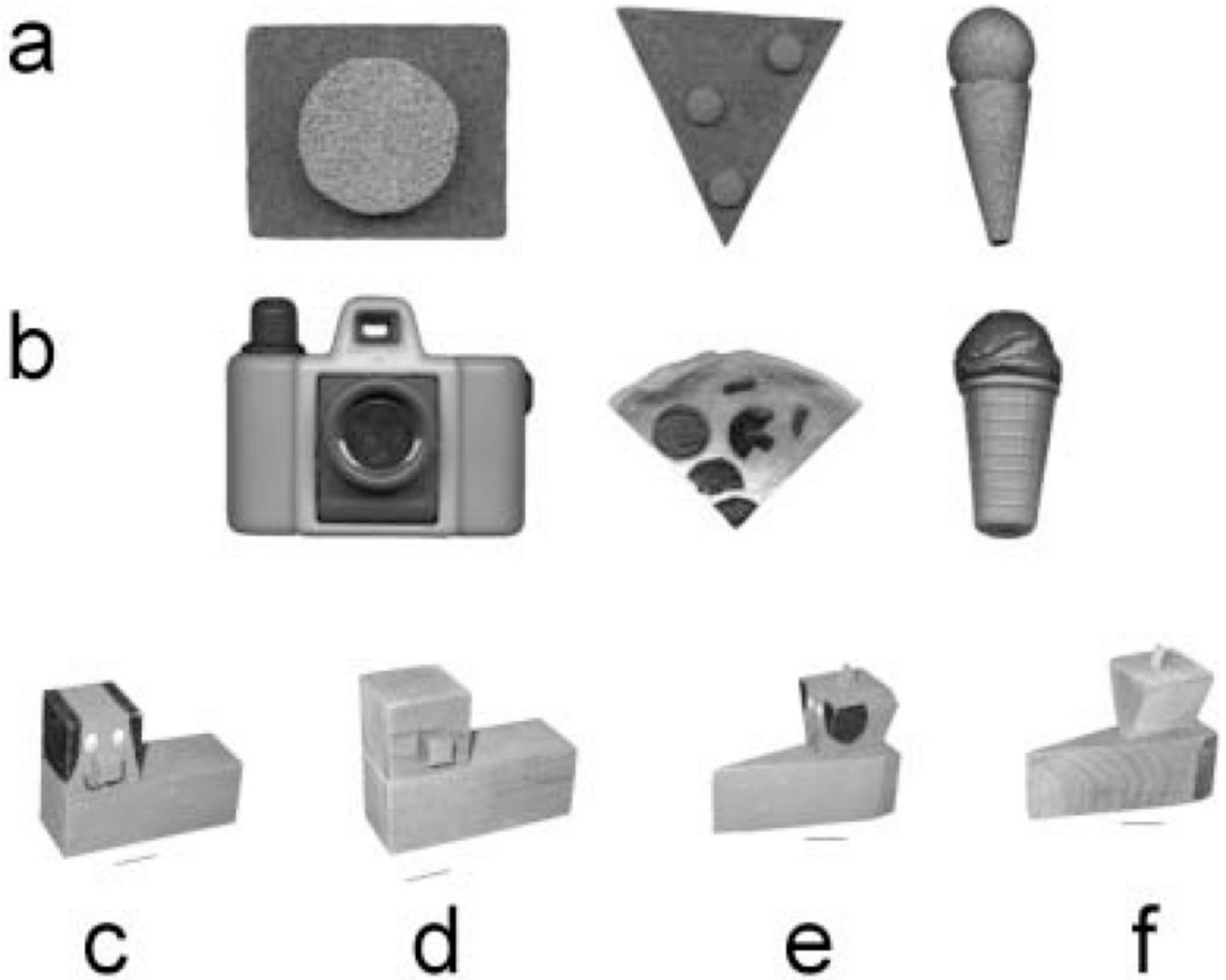
Fig. 2.
Stimuli used in experiments on shape caricature recognition: (a) sparse three-dimensional caricatures of the geometry of common categories, (b) richly detailed and lifelike instances, (c) caricatures with localized category specific features, (d) caricatures with no added features, (e) scrambled caricatures with localized category specific features, and (f) scrambled caricatures with no features.
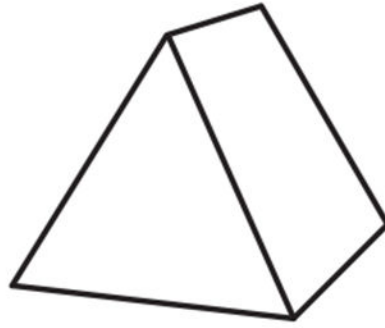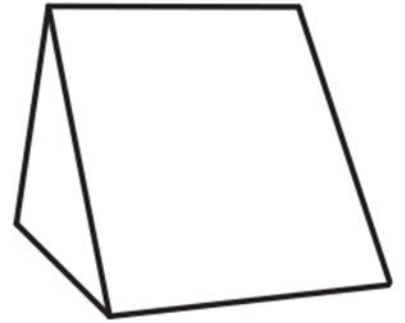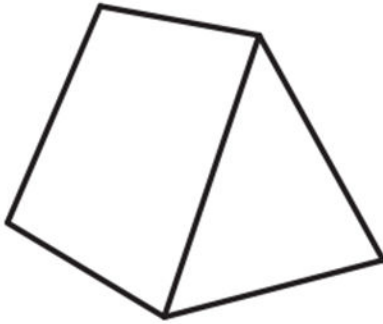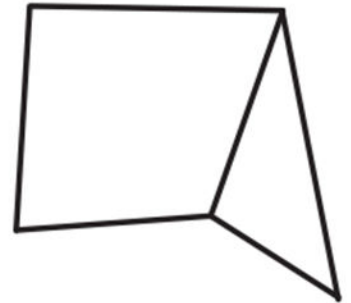
Fig. 3.
Rotational possibilities of a simple object. Given a view of an object from one side (a), adults expect to see views of volumes (such those shown in b) and not hollow shells (such as the views in c).