Journal of Cognition and Development

Young Children's Self-Generated Object Views and Object Recognition

Karin H. James [a] , Susan S. Jones [a] , Linda B. Smith [a] & Shelley N. Swain [a]

[a] Indiana University

Click for updates

PLEASE SCROLL DOWN FOR ARTICLE

ARTICLES

Young Children's Self-Generated Object Views and Object Recognition

Karin H. James, Susan S. Jones, Linda B. Smith, and Shelley N. Swain

Indiana University

Two important and related developments in children between 18 and 24 months of age are the rapid expansion of object name vocabularies and the emergence of an ability to recognize objects from sparse representations of their geometric shapes. In the same period, children also begin to show a preference for planar views (i.e., views of objects held perpendicular to the line of sight) of objects they manually explore. Are children's emerging view preferences somehow related to contemporary changes in object name vocabulary and object perception? Children aged 18 to 24 months old explored richly detailed toy objects while wearing a head camera that recorded their object views. Both children's vocabulary size and their success in recognizing sparse three-dimensional representations of the geometric shapes of objects were significantly related to their spontaneous choice of planar views of those objects during exploration. The results suggest important interdependencies among developmental changes in perception, action, word learning, and categorization in very young children.

Children routinely investigate objects in their environment through manual exploration. This exploration may be a crucial step in the construction of stored object representations at a very early age, as manual exploration allows children to encode multiple views of objects that may not be acquired through observation alone (Pereira, James, Jones, & Smith, 2010; Perone, Madole, Ross-Sheehy, Carey, & Oakes, 2008; Ruff, 1984; Soska, Adolph, & Johnson, 2010). A crucial question, then, is whether manual exploration actually facilitates object recognition—a reflection of stored object representations.

The ability to recognize common objects from a few geometric components is well established in mature visual object recognition (Biederman, 1995). Given caricatures composed of just two to four volumes in the proper relational structure, adults readily recognize instances of basic-level categories. Previous studies indicate that the ability to recognize well-known objects—a chair, a dog—from similarly sparse information about object shape first emerges

---

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

between the ages of 18 and 24 months old (Pereira & Smith, 2009; Smith, 2003). Using a name comprehension task, Smith examined 18- and 24-month-old children's ability to recognize three-dimensional (3D) object caricatures like those in Figure 1 as well as richly detailed instances of the same categories. Older children recognized the sparse geometric stimuli as well as they recognized rich instances of the same objects. Younger children recognized the rich instances but not the caricatures. Further studies have replicated this developmental trend and have also shown that: a) the ability to recognize sparse geometric representations is strongly correlated with productive vocabulary size—in fact, more strongly than with age (Pereira & Smith; Smith); b) late talkers show deficits in recognizing sparse caricatures (Jones & Smith, 2005); c) representation of sparse geometric structure supports broad generalization of categories (Son, Smith, & Goldstone, 2008); and finally d) recognition of geometric structure is more advanced for known object categories than for novel ones (Augustine, Jones, & Smith, 2011). All of these results suggest that changes in the representation of the geometric shapes of common objects occur between 1.5 and 2 years of age and that these changes are linked in some important way to learning the names of things.

Recently, Pereira et al. (2010) reported developmental changes across the same period in how toddlers hold objects during visual exploration: Older but not younger children held the objects to show themselves *planar* views. Planar views are defined as views in which: 1) the major axis of the object is approximately perpendicular or parallel to the line of sight, and 2) one axis is foreshortened (James, Humphrey, & Goodale, 2001; Perrett, Harries, & Looker, 1992). Older children's increasing preferences for planar views suggest an increasing sensitivity to the geometric structure of the objects. Further, adults engaged in visually exploring 3D objects in preparation for later object recognition also show a systematic preference for studying planar views (Harman, Humphrey, & Goodale, 1999; James et al.). Thus, the developmental change in how children hold and view objects could be related to the developmental change in object recognition: Both could reflect increasingly sparse and increasingly geometric representations of object shape. This possibility was tested in the present study, in which 18- to 24-month-old children first participated in visual and manual exploration of held objects and then in a test of
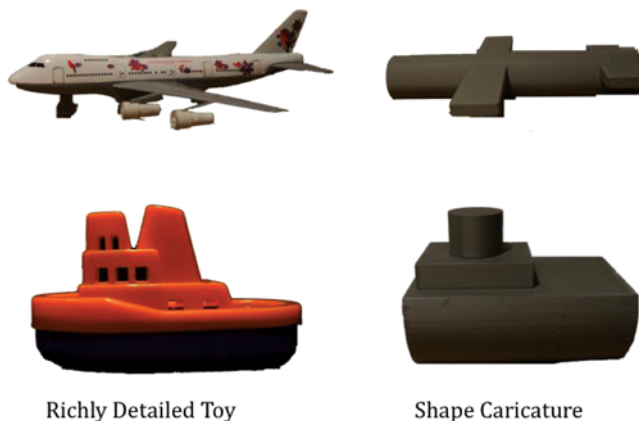

Richly Detailed Toy   Shape Caricature

FIGURE 1 Examples of stimuli: richly detailed toys and shape caricatures.

their recognition of sparse geometric versions of those objects. The key empirical question is whether the children's ability to recognize sparse geometric representations of common objects—as a measure of object recognition—is related to the views of objects that they showed to themselves—manual exploration. We also ask whether children's self-generated object views are related, like shape caricature recognition, to word learning, measured as productive vocabulary size.

METHOD

Participants

Participants were 36 children aged between 18 and 24 months old ($M_{age} = 1;9$, $SD = 2.2$ months) who were recruited from a working- and middle-class population in the Midwest. An additional 8 children were recruited but would not tolerate a head camera used to measure children's self-generated object views. Parents completed measures of language and general development: the MacArthur-Bates Communicative Development Inventory (MCDI; Fenson et al., 1994), a measure of productive vocabulary, and the Ages and Stages Questionnaires-Third Edition (ASQ-3; Squires & Bricker, 2009), a measure of social, emotional, and motor development.

Stimuli

Each child was tested on four target categories such that at least one tested category was an animal and at least one was a vehicle. The tested categories for each child were drawn from a total of 12 stimulus categories (airplane, bottle, chair, turtle, train, horse, boat, hammer, ice cream, pizza, car, and cup). For each child, 4 of these categories were the tested targets and the remaining 8 served as distractors. Half of the children were tested with airplane, bottle, chair, and turtle as the named targets and half were tested with train, horse, boat, and hammer as the named targets. The 2 distractors for each tested target were randomly selected without replacement to create four test trials in which no object was repeatedly seen by the child. Each category was represented by a richly detailed toy and by a shape caricature consisting of two to four plastic volumes representing the major parts of the object. The shape caricatures were generated by a 3D printer and were painted grey; examples are shown in Figure 1. The caricatures were designed to have the same volume as the rich instances they represented. Thus, the average dimensions of both the objects and their caricatures were 11.4 cm × 6.2 cm × 9.2 cm.

Children explored the rich version of each category in the object exploration task and were tested in a name comprehension task for recognition first of the caricatures and then of the rich objects. The named targets for each child were the same in both the caricature recognition and rich object recognition tasks.

Apparatus

A miniature video camera embedded in a headband and placed low on the forehead of the child recorded the child's self-generated object views. The camera was a WATEC model WAT-230A with 512 × 492 effective image frame pixels, weighing 30 g and measuring 36 mm × 30 mm × 30 mm. The lens used was a WATEC model 1920BC-5, with a focal length of f1.9 and an angle

of view of 115.2° on the horizontal and 83.7° on the vertical. Power and video cables were attached to the outside of the headband and were lightweight and long enough to provide freedom of movement to a seated participant.

To place the head camera on the child, one experimenter engaged the child in play while a second experimenter placed the head camera on the child's head in one swift movement. After placement on the head, the camera angle was adjusted so that when the child pushed a button on a toy, that button was in the center of the view. The child wore the head camera only for the object exploration task.

Tasks and Measurements

All children participated first in the object exploration task, then in the object recognition task with shape caricatures, then in the object recognition task with rich objects.

Object exploration task.   In this task, the child explored each of the rich and realistic toys representing the four target categories. The child sat on a chair removed from any surface so that each object had to be held to be explored. In each exploration trial, a single object was handed to the child in an orientation selected from 3×2 possibilities—oblique, major axis parallel to line of sight, major axis perpendicular to the line of sight—and each of these was realized with the object either upright or upside down. The experimenter encouraged the child to look at the object but never named it or directed attention to any of its parts. The four target objects were presented in an order that was randomly determined for each participant. An exploration trial for an object lasted until the child's interest in that object waned. Then the experimenter retrieved it and handed the child another object. When all objects had been offered to the child, the experimenter reoffered any objects that the child had visually explored for less than 15 s.

The head camera images were coded for duration of each exploration trial (determined from the number of frames) and for the object views. The object views during each trial were sampled at 1 Hz. The views were determined using a custom software application (Figure 2) that provided a 3D representation of each test object side by side with each frame from the video recording. The representation could be rotated to match the orientation of the object held by the child in each frame. The match yielded the 3D rotation of the test object in Euler angles (Kuipers, 2002; see Pereira et al., 2010, for details). From this output, a view was categorized as planar, using the procedure specified in Pereira et al. if the front, top, or side face of the object was within ±11.25° parallel or perpendicular to the child's line of sight. The planar views are a very small subset of distinct views that show only one side. No other views have this or any comparable distinguishing property (see Pereira et al. for more details). Using Monte Carlo simulations of random selection of views (see Pereira et al.), planar views are expected by chance alone to occur only 5.65% of the time.

Object recognition task.   On each trial of this task, children were presented with three objects and were asked to select the target named by the experimenter. Children were tested on the four caricature targets and then on the four rich targets. Two of the 36 children failed to complete all trials in the rich object recognition task. Prior to the experimental trials, a series of training trials using three common object categories (a banana, a cookie, and a cow) taught the
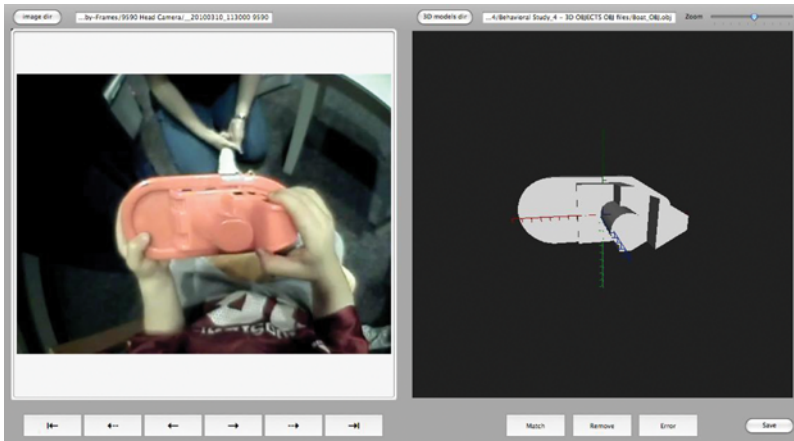
FIGURE 2 Screenshot of custom software program used to code children's object views in frames recorded by the head-mounted camera.

child to select the object named by the experimenter. These trials were repeated, with coaching, up to six times until the child selected the correct object when it was named.

For the experimental trials, first with shape caricatures and then with rich objects, the experimenter placed the target and two distractors 10 cm apart on a flat wooden tray (23 cm × 70 cm), aligned horizontally. While holding the tray out of reach of the child and looking into the child's eyes, the experimenter asked the child to select the named target (e.g., ''Where is the ___? Get me the ___.'') and then moved the tray to within the child's reach. If a child did not respond within 10 s, the tray was pulled back out of reach and the procedure was repeated. If there was still no response, the trial was repeated once at the end of the shape caricature recognition trials. Trial order was randomized across children. The experiment was video-recorded for later coding. The first object touched by the child among the three test objects in each trial was coded as the child's (correct or incorrect) response.

RESULTS

As a group, the children were slightly but significantly biased to look at planar object views more than predicted by chance: On average, 9.9% ($SD = 5.7\%$) of the coded frames showed planar views, exceeding the chance level of 5.65%, $t(37) = 4.70$, $p < .01$. Also, as a group, the children recognized an average of two or three of the shape caricatures ($n = 36$, $M = 2.75$, $SD = 1.2$) and about the same number of rich objects ($n = 34$, $M = 2.53$, $SD = 1.4$), $t(33) = 1.55$, $p = .13$. The children's scores in the shape caricature recognition task were not evenly distributed: 23 children recognized either three or four of the caricatures (i.e., they ''passed'' the shape caricature recognition task; $M = 3.56$, $SD = 0.51$) and 13 children recognized zero to two (i.e., they ''failed'' the shape caricature recognition task; $M = 1.38$, $SD = 0.77$). The central question for the study was whether the children who recognized most or all of the sparse geometric caricature shapes also showed themselves more planar views of objects they were exploring than did the children who recognized few or none of the caricatures.

As shown in Table 1, children who passed the shape caricature recognition task did look at significantly more planar object views on average than did children who failed that task. This is first evidence of a relation between the way that children hold objects for viewing and the way that they process the shapes of those objects.

As in Smith (2003), children who passed the shape caricature recognition task also had significantly larger productive vocabularies than did children who failed that task. Not surprisingly, children who recognized more shape caricatures and had more words also recognized more of the object names in the detail-rich object recognition task. What is more interesting is the finding that children with smaller vocabularies recognized fewer shape caricatures of objects even when they *did* know the object names. As shown in Table 1, the mean conditional probability that children would correctly identify a shape caricature, given that they had correctly identified its detail-rich counterpart, was significantly lower for the children who failed the shape caricature recognition task. It appears, then, that children who correctly identified more shape caricatures did so not just because they knew more object category names, but also because they more readily represented the geometric structures of known categories.

In short, children who knew more words gave evidence of greater sensitivity to the geometric structure of objects than did children who knew fewer words, and they did so in two ways—by recognizing more objects from sparse information about geometric shape alone and by showing the beginnings of the bias found in adults to hold objects in such a way as to show themselves planar views of the objects. Otherwise, the two groups of children did not differ in their average general levels of development (ASQ-3 scores, Passed, $M = 241.7$, $SD = 24.6$; Failed, $M = 232.3$, $SD = 42.9$), $t(34) = 0.84$, $p = ns$. Nor did the two groups differ in average age (Passed, $M = 21.2$ months, $SD = 2.09$; Failed: $M = 20.2$ months, $SD = 2.2$), $t(34) = 1.65$, $p = .11$, or in the amount of time they spent exploring the detail-rich objects (Passed, $M = 129.2$ frames, $SD = 68.3$; Failed, $M = 108.5$ frames, $SD = 69.0$), $t(34) = 0.87$, $p = .39$.

TABLE 1
Results of *t*-Tests Comparing Means of Children With High Shape Caricature Recognition Scores ($n = 23$) and Children With Low Shape Caricature Recognition Scores ($n = 13$)

| | Shape caricature recognition | | | | | |
| | "Passed" | | "Failed" | | | |
	Mean	SD	Mean	SD	t	p value
Planar Views (# frames)	14.04	9.3	8.08	5.5	2.11	.04
Total Words MacArthur-Bates Communicative Inventory (MCDI)	254	145	144	157	2.12	.04
Rich Object Recognition	3.27	0.88	1.17	1.19	5.87	.001
p (Shape Car. Rich Obj. Recog.)	0.96	0.10	0.56	0.50	3.62	.001*

Note. Table includes number of frames showing planar views; total productive vocabulary size; rich object recognition scores; shape caricature recognition scores; and the probability of recognizing a shape caricature given that the rich object from the same category was recognized.

All $df = 34$ except *$df = 28$.

DISCUSSION

Human visual object recognition is fundamental to all aspects of human cognition. Thus, developmental changes in the higher-level visual processes through which objects are recognized might be expected to have consequences for a range of cognitive functions. Previous studies have shown that children between the ages of 18 and 24 months old are increasingly able to recognize common objects when category-specific diagnostic features are removed and only sparse information about geometric shape remains. Children's success in recognizing common objects from such sparse shape information is strongly related to the number of nouns in their productive vocabularies. This suggests that changes in visual object recognition are linked in some way to object name learning.

The present results extend these findings by showing that the increases in children's object recognition from sparse shape information are also related to how children spontaneously hold objects when visually exploring those objects. One possible explanation for this relation is suggested by two considerations. Unified representations of 3D shape require that multiple views of an object be integrated (e.g., Farivar, 2009; Graf, 2006). Virtually all contemporary theories of how such object-centered representations are built propose that the object's major axis determines the frame of reference for that integration. Planar views are views that align that object-centered frame of reference either perpendicular or parallel to the line of sight. Thus, how the child holds and rotates objects for viewing is both likely to be a critical source of information for this integration and likely to influence the efficiency of that integration. In short, it may be that children with a preference for planar views score well on a test of sparse object recognition because they are guided by their view preferences that possibly support the formation of more coherent representations of object shape.

The present results also indicate that children's preference for planar views during visual exploration is related to vocabulary size. Theories (as well as studies with nonhuman primates) suggest that the construction of sparse shape representations is critically dependent on category learning (Doumas & Hummel, 2010; Edelman & Intrator, 2003; Kourtzi & Connor, 2011). Thus, it could be that both children's dynamic viewing preferences and their ability to represent common objects in terms of sparse geometric shape depend on category learning, which is indexed by vocabulary size. Figure 3 summarizes one hypothesis about the causal relations among preferential viewing, caricature recognition, and vocabulary. The hypothesis is that planar views—that is, views organized around the major axes—promote the formation of whole-object representations of specific instances (Pereira et al., 2010). Object name learning, as it teaches which multiple instances are in the same category, also promotes the extraction of the structural shape properties common to those multiple instances—multiple whole-object representations of chairs, for example—to build a sparse representation common to all kinds of chairs. Thus, as illustrated in the figure, the correlation between learning object names and a preference for planar views may be indirect and emerge because they are both related to the building of these sparse representations. However, it is also possible that there are more direct links. Having whole-object representations at either the instance or category level may directly benefit the learning of object names. In addition, knowing many object names and thus having developed sparse whole-object representations of many object categories may further the likelihood of holding even novel objects in such a way as to oversample planar views.

These possibilities are both plausible and consistent with existing research. However, because the directions of the relations discovered in the present study are not known, it is also possible that children's viewing preferences and their vocabulary development (with the category learning it implies) are distinct developments that support building 3D representations of object
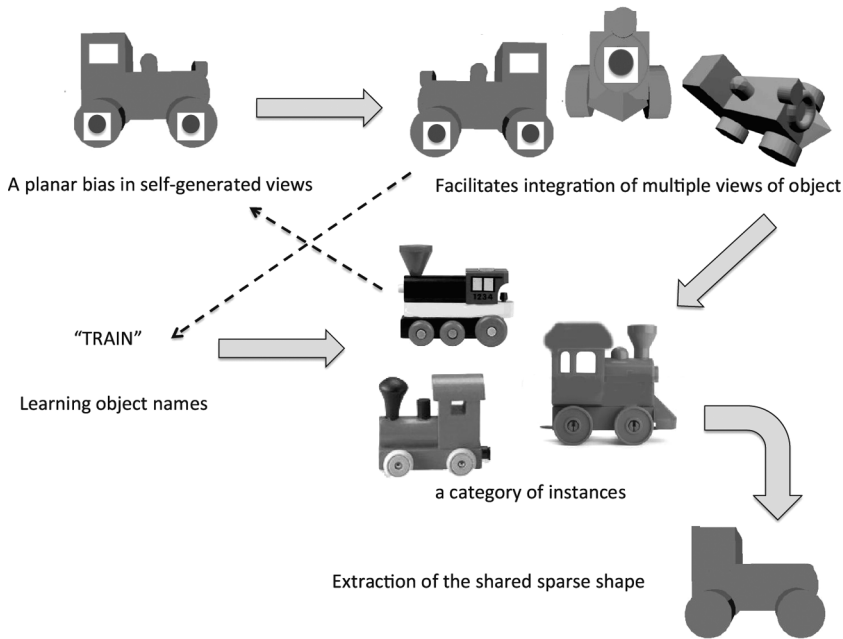
FIGURE 3 Illustration of a set of hypotheses about how a preference for planar views, learning object names, and shape caricature recognition are related. See text for discussion.

shape; or that both depend on prior developments in forming abstract shape representations. The question of how these three developmental changes are actually related is an important one for future research. And it is a question best answered not with correlational evidence, but through epigenetic studies that provide children with the hypothetically critical experiences.

In sum, growing evidence—now in a variety of different tasks—suggests that significant changes in the development of sparse representations of 3D object shape occur during the same period in which children learn object names and in which their functional actions on and manual explorations of objects also increase. These changes in object recognition are linked to lexical learning in two ways: They are correlated with vocabulary size in typically developing children, and they are late-emerging and fragile in children with language delay. Previous work has shown that these changes in object recognition are also linked to children's functional play with objects (Smith & Jones, 2011), and here we further show that they are linked to preferences in how children hold and view objects. Although details are still lacking, these results provide a sketch of a developing system in which changes in perception, action, categorization, and word learning all interact, and they suggest that this system provides a useful context for understanding developmental process in all of these domains.

FUNDING

REFERENCES

Augustine, E., Jones, S., & Smith, L. B. (2011). Parts and relations in young children's shape-based object recognition. *Journal of Cognition and Development*, *12*, 556–572.

Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn & D. N. Osherson (Eds.), *Visual cognition: An invitation to cognitive science* (2nd ed., Vol. 2, pp. 121–165). Cambridge, MA: The MIT Press.

Doumas, L. A. A., & Hummel, J. E. (2010). A computational account of the development of the generalization of shape information. *Cognitive Science*, *34*, 698–712.

Edelman, S., & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, *27*, 73–109.

Farivar, R. (2009). Dorsal–ventral integration in object recognition. *Brain Research Reviews*, *61*, 144–153.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*, 1–173.

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, *132*, 920–945.

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*, 1315–1318.

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: Where the action is. *Canadian Journal of Experimental Psychology*, *55*, 111–120.

Jones, S. S., & Smith, L. B. (2005). Object name learning and object perception: A deficit in late talkers. *Journal of Child Language*, *32*, 223–240.

Kourtzi, Z., & Connor, C. E. (2011). Neural representations for object perception: Structure, category, and adaptive coding. *Annual Review of Neuroscience*, *34*, 45–67.

Kuipers, J. B. (2002). *Quaternions and rotation sequences: A primer with applications to orbits, aerospace, and virtual reality*. Princeton, NJ: Princeton University Press.

Pereira, A. F., James, K. H., Jones, S. S., & Smith, L. B. (2010). Early biases and developmental changes in self-generated object views. *Journal of Vision*, *10*, 1–13.

Pereira, A. F., & Smith, L. B. (2009). Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental Science*, *12*, 67–83.

Perone, S., Madole, K. L., Ross-Sheehy, S., Carey, M., & Oakes, L. M. (2008). The relation between infants' activity with objects and attention to object appearance. *Developmental Psychology*, *44*, 1242–1248.

Perrett, D. I., Harries, M. H., & Looker, S. (1992). Use of preferential inspection to define the viewing sphere and characteristic views of an arbitrary machined tool part. *Perception*, *21*, 497–515.

Ruff, H. (1984). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology*, *20*, 9–20.

Smith, L. B. (2003). Learning to recognize objects. *Psychological Science*, *14*, 244–250.

Smith, L. B., & Jones, S. (2011). Symbolic play connects to language through visual object recognition. *Developmental Science*, *14*, 1142–1149.

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*, 626–638.

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*, *46*, 129–138.

Squires, J., & Bricker, D. (2009). *Ages and Stages Questionnaires: A parent-completed, child-monitoring system* (3rd ed.). Baltimore, MD: Paul H. Brookes Publishing Company.