

Early biases and developmental changes in self-generated object views

Alfredo F. Pereira

Psychological and Brain Sciences, Indiana University,
Bloomington, IN, USA



Karin H. James

Psychological and Brain Sciences, Indiana University,
Bloomington, IN, USA



Susan S. Jones

Psychological and Brain Sciences, Indiana University,
Bloomington, IN, USA



Linda B. Smith

Psychological and Brain Sciences, Indiana University,
Bloomington, IN, USA



Object recognition depends on the seen views of objects. These views depend in part on the perceivers' own actions as they select and show object views to themselves. The self-selection of object views from manual exploration of objects during infancy and childhood may be particularly informative about the human object recognition system and its development. Here, we report for the first time on the structure of object views generated by 12 to 36 month old children ($N = 54$) and for comparison adults ($N = 17$) during manual and visual exploration of objects. Object views were recorded via a tiny video camera placed low on the participant's forehead. The findings indicate two viewing biases that grow rapidly in the first three years: a bias for planar views and for views of objects in an upright position. These biases also strongly characterize adult viewing. We discuss the implications of these findings for a developmentally complete theory of object recognition.

Keywords: visual object recognition, active object inspection, viewpoint selection

Citation: 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(11):22, 1–13, <http://www.journalofvision.org/content/10/11/22>, doi:10.1167/10.11.22.

Introduction

Most theories of visual object recognition emphasize the importance of the specific object views that a perceiver has experienced (Peissig & Tarr, 2007). From this perspective, the object views experienced by young children would seem especially important as they are the foundational experiences for the object recognition system. Object views are tightly dependent on perceivers' own actions, and in particular, on how they hold and what they do with objects. For this reason alone, it seems likely that the statistical properties of object views will change with development (Smith, 2009; Soska, Adolph, & Johnson, 2010). However, we know almost nothing about the object views that characterize early experience. Here, we present the first developmental evidence on the object views generated by 12 to 36 month old children as they manipulate and explore objects. We focus on this developmental period because it is one in which children's manual interactions with objects become increasingly sophisticated (Ornkloo & von Hofsten, 2007; Ruff & Saltarelli, 1993; Ruff, Saltarelli, Capozzoli, & Dubiner, 1992; Street, James, Jones, & Smith, *in press*); because it

is the period in which children learn common object categories (Smith, 1999); and because it is also a period of significant change in visual object recognition (Smith, 2009). We specifically ask whether and when two biases that characterize adult object views are evident in the object views of very young children.

When adults self-generate object views for the purpose of later recognizing the viewed objects, they dwell on or near the planar sides—preferring, for example, views near the view shown on the left panel in [Figure 1](#) over views near the view shown on the middle panel (Harries, Perrett, & Lavender, 1991; James et al., 2002; Perrett & Harries, 1988; Perrett, Harries, & Looker, 1992). More precisely, a planar view is defined as a view 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. Further, if an object has flat surfaces, adults prefer to view the object so that the flat surface is perpendicular to the line of sight (Harman, Humphrey, & Goodale, 1999; Harries et al., 1991; James, Humphrey, & Goodale, 2001; James et al., 2002; Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008; Locher, Vos, Stappers, & Overbeeke, 2000; Niemann, Lappe, & Hoffmann, 1996; Perrett & Harries, 1988; Perrett et al.,



Figure 1. A planar view (left panel) and a non-planar view (middle panel). The right panel shows the coordinate system used to define the object view orientation with respect to the viewer. A viewpoint is considered a planar view if the major axis is approximately perpendicular or parallel to the line of sight.

1992). Some studies suggest that subsequent object recognition benefits from object views that conform to these biases (James et al., 2001), a result that implies that these views may build better object representations. However, adults' preferences for planar views are specific to the dynamic viewing of 3-dimensional objects; given pictures, adults prefer $3/4$ views (Blanz, Tarr, & Bülthoff, 1999; Boutsen, Lamberts, & Verfaillie, 1998; Palmer, Rosch, & Chase, 1981). The specific mechanisms that underlie dynamic view selection and the relevance of those mechanisms to object representation are not as yet well understood. In the [General discussion](#), we consider several possibilities in light of the developmental data.

Adults are also biased to view objects in a standard vertical orientation (Blanz et al., 1999; Harman et al., 1999) and show this “upright” bias for appropriately structured novel objects (objects with a clear main axis of elongation and upright direction) as well as for familiar objects with a canonical orientation (Blanz et al., 1999; James et al., 2001; Perrett et al., 1992). This orientation bias characterizes both picture and dynamic viewing and has been assumed to play a role in maintaining a stable frame of reference for integrating multiple views of a single object and for comparing a view to representations in memory (Marr & Nishihara, 1978; Rock, 1973). Preferences for viewing objects in upright orientations have been documented in older preschool and school-aged children with respect to pictures (Landau, Hoffman, & Kurz, 2006; Picard & Durand, 2005). However, there are no studies examining this bias in the 3-dimensional views generated by children through their own actions and no studies of this potential bias early in development.

We ask: What are the developmental origins of these viewing biases? If the biases are developmentally early, they would constitute a strong constraint on experienced views and object representations. If, however, the biases that characterize adults' self selected views are not present early in development but emerge late in the development

of object recognition, then they might be better understood as end products rather than formative constraints. The timing of developmental changes in these biases may also be revealing as to the responsible mechanisms and to their possible role in building object representations. Pertinent to this last point, a growing number of studies suggest fundamental changes in how children recognize common objects during the period between 12 and 36 months. In particular, early in this developmental period, children appear to recognize common objects via piecemeal features (Pereira & Smith, 2009; Rakison & Butterworth, 1998; Smith, 2009), in a manner similar to that proposed by Ullman's fragment model of object recognition (Ullman, 2007; Ullman, Vidal-Naquet, & Sali, 2002). But by the end of this developmental period, children rely more on the geometric properties of whole-object shape (Pereira & Smith, 2009; Smith, 2003), in a manner similar to that proposed by Hummel and Biederman (Biederman, 1987; Hummel & Biederman, 1992). These developmental changes in object recognition are also strongly related to children's knowledge of object categories (Smith, 2003).

The present study examines young children's self-generated object views during this developmental period. Are these views biased in the same ways as adults' self-generated views? Do these biases change during this developmental period? [Experiment 1](#) is the main developmental study; [Experiment 2](#) reports data from adults in the same task used with children.

Experiment 1

Methods

Participants

The participants were 54 children (25 female) in four age groups (labeled according to the youngest age in the group): 1 year olds (N = 14, range 12 to 17 months,

$M = 15.4$), 1½ year olds ($N = 14$, range 18 to 23 months, $M = 22.0$), 2 year olds ($N = 14$, range 24 to 29 months, $M = 27.5$), and 2½ year olds ($N = 12$, range 30 to 36 months, $M = 34.7$). Twenty-two additional children were recruited but did not contribute data due to failure to tolerate the measuring equipment or because of equipment malfunctions.

Stimuli

To ensure the generality of the results, the stimuli were constructed to represent a wide variety of shapes of common objects. Specifically, the stimuli were derived from 8 common object categories and were created in two versions: (1) a typical toy version painted a uniform color and (2) a variant created to maintain the geometric structure of the original (major axes and major parts and concavities) and painted a uniform color as shown in [Figure 2](#). Each child saw a random selection of 4 toy and 4 simplified versions with the constraint that no child saw the toy and geometric variant from the same originating category. A hypothetical box bounding each object had an average volume of $M = 595 \text{ cm}^3$ (range 144 to 976 cm^3 , $SD = 344 \text{ cm}^3$). Objects had an average weight of 72.2 g (range 20 to 160 g, $SD = 38.4 \text{ g}$).

Apparatus

The child's self-generated object views were captured by a mini video camera mounted on a headband, a method used in a number of recent studies of active vision (Aslin, 2009; Smith, Yu, & Pereira, 2010; Street et al., *in press*; Yoshida & Smith, 2008). Prior studies indicate that when toddler's are actively holding and viewing objects, their eye and head directions are typically aligned (with eyes leading heads in shifts of direction by about 500 msec) as shown by frame-by-frame analyses of eye-gaze direction—measured by a human coder using a third-person perspective camera—and head camera direction (Yoshida & Smith, 2008) and as recently confirmed by measures of head motion (via motion sensors) and eye gaze direction (via remote eye tracking) (Shen, Baker, Candy, Yu, & Smith, *in preparation*).

As shown in [Figure 3](#), the head-camera is mounted on a rubber strip, measuring $58.4 \text{ cm} \times 2.54 \text{ cm}$. The band was fit onto the child's head and fastened at the end with Velcro. The camera could be rotated slightly to adjust viewing angle. The camera was a WATEC model WAT-230A with 512×492 effective image frame pixels, weighing 30 g and measuring $36 \text{ mm} \times 30 \text{ mm} \times 30 \text{ mm}$. 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; they were lightweight and long enough to not limit movement while the participant was seated. In total, the equipment

weighed 100 g. A second camera was positioned facing the child to provide a view of the child's face and the object manipulations.

A video mixer (model ROLAND Edirol V-4) synchronized the video signal from the head camera, the video signal from the camera facing the child and the audio signal coming from an omnidirectional microphone placed in the room. The output of the video mixer was recorded in a digital video file with a standard video capture card at 25 Hz. For coding, image frames were sampled at the rate of 1 Hz.

Procedure

The child sat on a small chair and the parent sat nearby. A removable table with a push-button pop-up toy was placed close to the child to assist in placement of the head camera. One experimenter directed the child's attention to the pop-up toy while a second experimenter placed the head camera on the child's head. This was done in one movement. The parent assisted by placing his/her hands above the child's hands, gently preventing any movement of the hands to the child's head as the head camera was put on. With the child engaged in pushing buttons, the head camera was adjusted so that when the child was pushing a button, that button was in the center of the head-camera field. The table was then removed so that when the child was given a toy, there was no support for the toy other than the child's own hands.

Children were given eight objects one at a time in random order. Children were allowed to examine each object for as long as they wished. When the child's interest waned, the experimenter took back that object and gave the child the next one. Objects were handed to the child in an initial orientation that was randomly determined for each object and child.

Coding

Only those frames recorded while the child was both holding and looking at an object were coded. The second camera directed at the child's eyes was used to ensure that the child's eye gaze was directed at the object for each coded head camera image. To measure object orientation a human coder found the closest match between a head-camera image frame matched and a three-dimensional model of the object. This was done via a custom-made software application shown in [Figure 3](#). This software allowed the coder to make side-by-side comparisons of images taken from the head-mounted camera and images of a computer-rendered model of the same object and to rotate the computer model to match the view in head-camera image. The rotation of the three-dimensional object was specified with Euler angles, and the standard used was YXZ or heading-attitude-bank (Kuipers, 2002). The coding application used the following coordinate



Figure 2. The 16 stimulus objects: the original toy object, and the geometrically simpler variant. The line close to each object is one inch in length. The filled arrow points to the defined top of the object and the dashed arrow to the front. The axis positive direction is in the inverse direction to the depicted arrow.

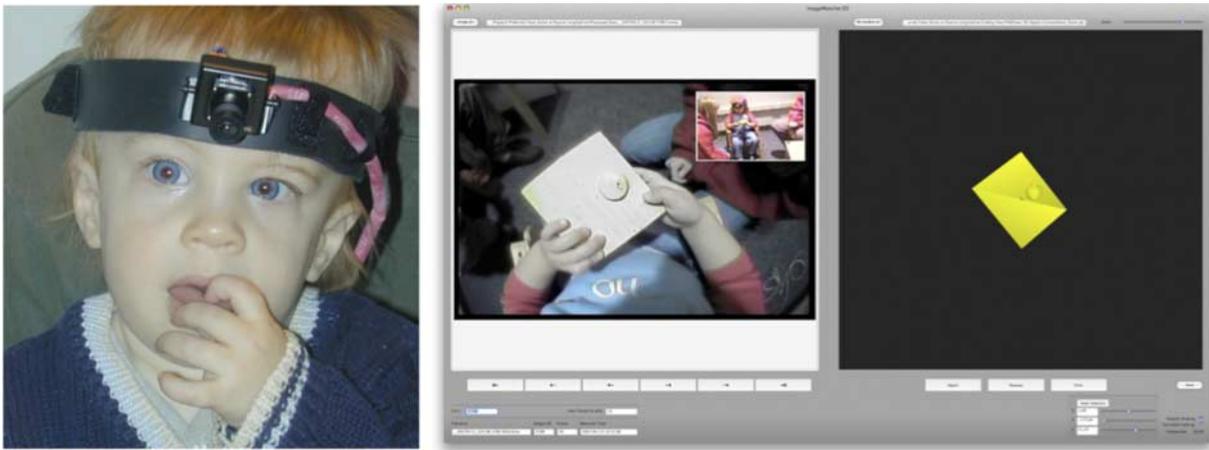


Figure 3. A subject wearing the head-mounted camera (left) and snapshot of the coding application used to measure object orientation (right). In the left side of the application the human coder sees image frames from the head-mounted camera video; on the right a three-dimensional model of the current object in view can be rotated until it matches the current frame.

system: XY was in the picture plane, with X positive towards the right and Y positive in the up direction; the viewer's line of sight was Z, with the positive direction away from the viewer. Figure 2 shows the X-axis and the Y-axis of each object. Coders also marked the onset and offset of every trial in order to measure how long each object was held. The first frame in which an object could be seen in the head-mounted camera image, and orientation could be measured, was coded as the onset; the last frame in which the child was independently holding and looking at the object was coded as the offset.

The orientation of the held object with respect to the picture plane's upright was determined by calculating the angle between the three-dimensional vector perpendicular to the object's top and the unit vector pointing up, that is, to vertical axis of the picture plane. When this angle was less than 90 deg the object view was coded as upright. The canonical top for each of the toy objects (with the exception of the shovel) was defined in terms of the functional upright of the original toy object and is indicated in Figure 2. This definition of upright was used for both versions of an object category. For the one object without an obvious upright (the shovel), the object was positioned on the horizontal plane and the main axis of elongation defined the vector normal to the front side plane; then the axis perpendicular to the main axis of elongation and parallel to the horizontal plane defined the sides of the object. The top and front planes of each object as indicated in Figure 2 were used to combine data across objects for analyses. The two vectors marking the top and the front side define a bounding box for the object; the faces of this box are always parallel to one of object's sides. When the object has curved surfaces, the faces are either perpendicular or parallel to the object's principal axis. The sides of each bounding box are thus the six planar views, front, back, top, bottom and sides.

Results and discussion

Both of children's hands were in contact with the object on more than 80% of the frames, which is consistent with other studies of manual holding and exploration in young children (Bushnell & Boudreau, 1993; Ruff et al., 1992). Children were free to explore the objects as long as they wanted and older children typically did so longer than younger children, $F(3, 50) = 2.99, p < .05$; the mean holding times were 15.4, 22.0, 27.5, and 34.7 seconds, for the four age groups respectively. Preliminary analyses of all dependent variables with respect to gender led to no reliable effects; accordingly this factor is not discussed further.

Dwell time maps

Dwell time maps provide a comprehensive way to visualize the developmental changes in self-generated object views. To construct these maps, the object orientation, expressed as three Euler angles, was transformed into azimuth, elevation, and tilt coordinates. Tilt corresponded to rotations about the line of sight and azimuth and elevation corresponded to the coordinates in the object's viewing sphere (see Figure 1). If one imagines moving on the surface of a sphere centered on the object and looking in the direction of the sphere's center, visiting all points on the sphere will produce all possible object views, under the simplification that rotations about the line of sight produce the same view. The spherical coordinates of any point on the sphere (azimuth and elevation) therefore uniquely identify each object view. For the maps shown in Figure 4, the azimuth and elevation angles were binned as follows: the azimuth interval of $[-180 \text{ deg}; +180 \text{ deg}]$ was divided into 16 equal bins of 22.5 deg. Bins were centered at the points -150 deg to

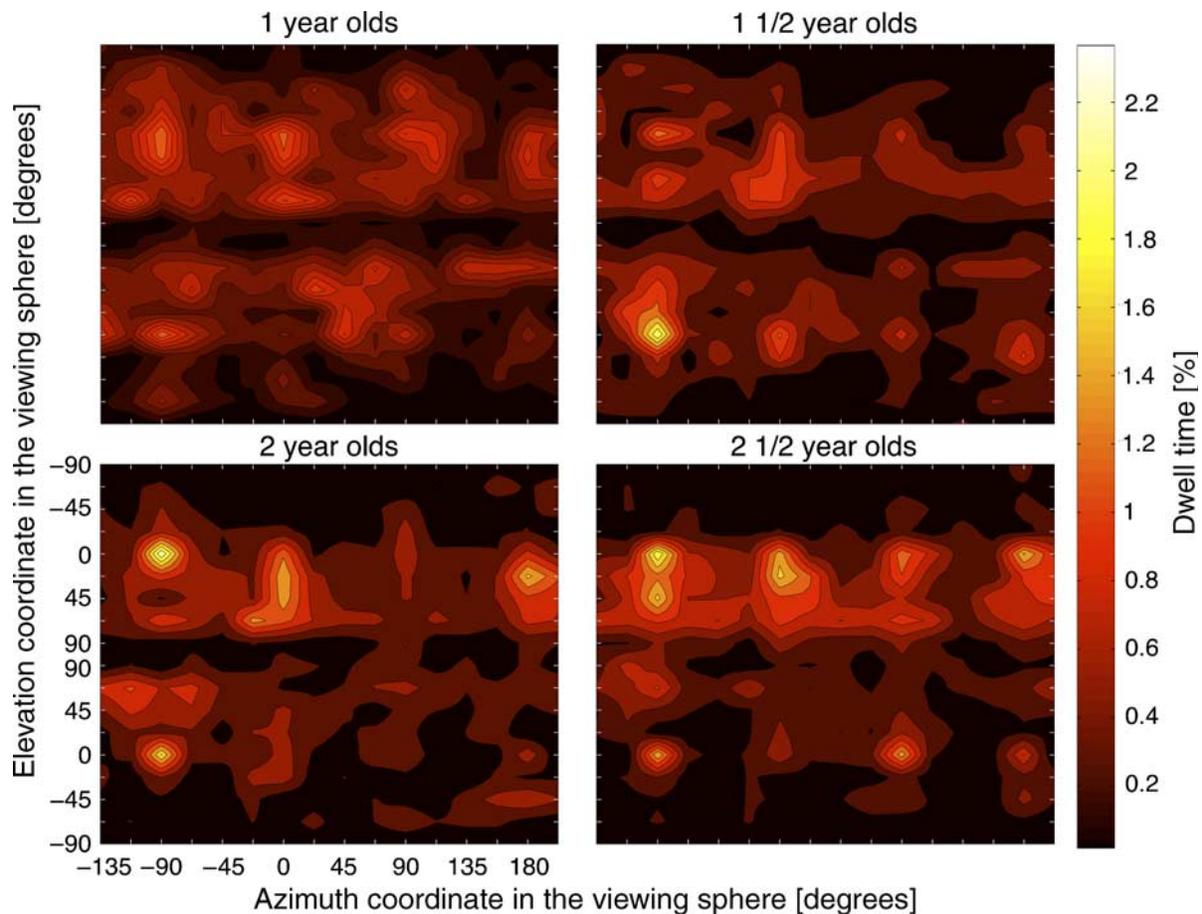


Figure 4. Dwell time map across all objects for the four age groups in [Experiment 1](#). [Figure 1](#)'s right panel depicts the coordinate system used to define object orientation with respect to the line of sight. Each map is shown with the elevation axis duplicated; the top half corresponds to the object in an upright orientation and the bottom half shows the reverse. The order of the elevation values in this axis is arranged so that the object view is symmetrical about the map's horizontal mid-line. The azimuth and elevation coordinates for the six planar views are as follows: $+90^\circ$ elevation, top; -90° elevation, bottom; $(-90^\circ, 0^\circ)$, front; $(0^\circ, 0^\circ)$, side 1; $(+90^\circ, 0^\circ)$, back; and $(+180^\circ, 0^\circ)$, side 2.

$+180$ deg in increments of 22.5 deg (the first half-bin was added to the last half-bin). The elevation interval of $[-90$ deg; $+90$ deg] was divided using the same bin size, but the first and last bin were 11.25 deg in width, generating nine bins in total.

This particular slicing of the viewing sphere produced three types of viewpoint bins, each defined by the type of object view at its center: a planar view, a $3/4$ view (with both of these views defined as in the categorical analysis), and the views between these planar and $3/4$ views. We normalized the resulting 2D matrix by the total number of frames (i.e. it was expressed as proportion of total dwell time.) In addition, and to maintain some information about the original orientation in the picture plane, the dwell time maps are shown with the elevation axis duplicated; the top part corresponds to the object in an upright orientation and the bottom part shows the views when the object is upside-down. Finally, for plotting, the matrix was interpolated and contour lines were superimposed.

The major result is immediately apparent in [Figure 4](#): Younger children's views are more disorganized—with views more distributed across the viewing sphere—than are older children's views. The oldest children's views, in contrast, are highly selective with views concentrated in the upper half of the plot, indicating that these objects were held in an upright orientation, and with views concentrated around the planar sides of the objects. These maps, constructed across a diverse set of objects, indicate a generalized and marked developmental progress toward a constrained sampling of object views. However, even in the youngest children's maps, there are hints of structure with increased frequencies of views broadly in the regions of the planar sides.

Planar versus $3/4$ views

The above conclusions about developmental changes in preferences for planar views were supported by statistical

comparisons of two broad categories of planar versus $3/4$ views. An object's orientation in the head-camera view was categorized as a planar view if one of the three signed angles between the Line of Sight (LoS) and the normal for front, top or side faces of the bounding box was within one of these intervals: 0 ± 11.25 deg; or 180 ± 11.25 deg. A viewpoint was categorized as a $3/4$ view if at least two sides of the bounding box were in view and the angle of LoS with both of them was inside the interval 45 ± 11.25 deg or a rotation of 90 deg on this interval (i.e. centered at 135 deg, 225 deg or 315 deg). A face of the object was considered to be in view if the unsigned (absolute) angle between LoS and the face normal was outside the interval 90 ± 11.25 deg. Broad categories are appropriate for this comparison because prior research indicates that adult participants do not present themselves with perfectly flat surfaces (Blanz et al., 1999; James et al., 2001; Perrett et al., 1992) and because *a priori* it seems likely that children's preferences would be less precise than adults. Monte Carlo simulations or random (that is unbiased) viewing yield expected values of 5.6% planar views and 33.1% three-quarter views.

As shown in Figure 5, this analysis indicates confirms the developmental changes in self-generated object views seen in the dwell time maps, with planar views increasing with age and $3/4$ views decreasing with age. A 2 (Object Version) \times 2 (View Type) \times 4 (Age Group) analysis of

variance for a mixed design which yielded only a significant interaction of View Type with Age Group, $F(3, 50) = 5.103$, $p < 0.005$. Older children spent more time on planar views and less time on $3/4$ views than did younger children. However, children in all age groups spend reliably more time on planar views than expected by random rotations (two-tailed t -test against a proportion of 5.6% from random inspection, $p < 0.005$, with Bonferroni correction for four comparisons). Only children in the oldest two age groups spent significantly *less* time on $3/4$ views than expected by the random distribution of views (two-tailed t -test against a proportion of 33.1% from random inspection, $p < 0.05$, with Bonferroni correction for four comparisons). Together these results suggest that even 1 year olds self-generated views are biased toward planar views but, and as apparent in the dwell time maps, the views of very young children are much less exact, noisier, than those of older children.

Figure 6 shows that these age-related differences in planar views characterize individual performance as a function of age. Age and time spent on planar views were positively correlated, $r(52) = 0.374$, $p < 0.01$, and age and time spent on $3/4$ views were negatively correlated, $r(52) = -0.327$, $p < 0.05$. These age-related effects are also not due to the age-related differences in the amount of time spent manually and visually exploring the objects; there was no correlation between total time spent manipulating

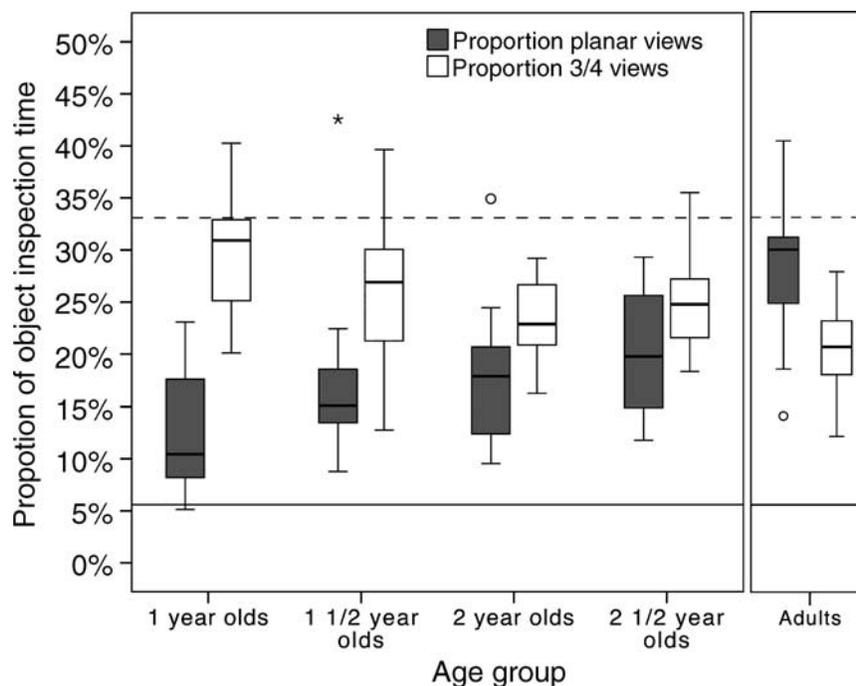


Figure 5. (Left panel) Experiment 1: box-and-whisker plot of relative time spent on planar views and on three-quarter views across all object manipulation time in four equal sized age groups: 1 year olds, $1\frac{1}{2}$ year olds, 2 year olds, $2\frac{1}{2}$ year olds. (Right panel) Experiment 2: box-and-whisker plot of relative time spent on planar views and on three-quarter views across all object manipulation time for adult participants. The proportion of time on planar views, expected by random viewing, is given by the solid reference line, and the equivalent proportion on three-quarter views, is given by the dashed reference line.

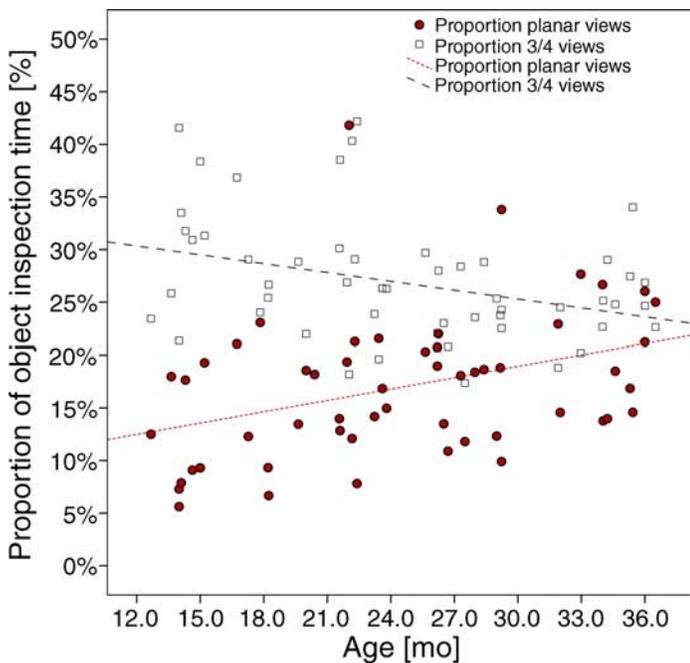


Figure 6. Proportion of object manipulation time where the viewpoint was a planar or a three-quarter view (solid circles or empty squares respectively) by participant's age. The regressions of age on the proportion of time spent looking at planar views and at three-quarter views on age are shown separately.

objects and time spent viewing planar views, $r(52) = 0.12$, $p > 0.1$, or time spent on three-quarter views, $r(52) = -0.209$, $p > 0.1$.

The findings provide two new insights relevant to early object recognition. First, the object views that very young children show themselves become more biased and more centered on planar views. Second, the youngest children showed themselves planar views more frequently than expected by random object rotations and thus even one-year olds' self-generated object views are biased (albeit weakly).

Object orientation with respect to the upright

The dwell map analysis also indicates a second developing bias. Upright orientation for the dwell time maps was defined liberally (a view was considered upright if the angle between the normal vector to the object's topside and the unit vector pointing up in the picture plane was less than 90 deg) yet only the older children's dwell maps indicate consistent viewing of objects in upright orientations. A 4 (Age Group) analysis of variance of proportion of views with the objects in an upright location yielded a main effect of age, $F(3, 50) = 3.678$, $p < 0.02$, which is due to the significantly greater proportion of upright object views for the oldest children ($M = .64$, $SD = .07$) with respect to the three younger age groups ($M = .56$, $SD = .11$; $M = .50$, $SD = .15$; $M = .58$, $SD = .12$ for

the three younger groups). Thus, only children in the oldest age group spent significantly more time on upright views than expected by random (two-tailed t -test against a proportion of 50% from random inspection, $p < 0.001$ for the oldest age group, $p > 0.05$ for the remaining age groups, with Bonferroni correction for four comparisons). The bias for upright orientations is thus later than that for the planar views and perhaps much more dependent on specific experiences with the canonical orientations of objects.

Discussion

In summary, there are three main findings: First, there are significant changes in the views that children present to themselves during this developmental period, a period already marked as relevant to the development of human object recognition (Smith, 2009). Second, children's self-generated views becomes increasingly structured and biased in two ways, an increasing preference (and precision) for views around the planar sides of objects and an increasing preference for particular alignments of the objects with respect to upright. Third, there is structure in the self-generated views of even one-year olds who show a weak preference for the planar views that dominate older children's self-generated views. There is an additional organizational property in the dwell time maps that may also be worthy of note and that is also evident in even the youngest children's viewing. Along with heightened planar views, there is a band of dwell time activity connecting the 4 sides that are rotations around a single axis. Children are biased toward views around one axis of the object. This is evident at all ages although weaker for younger than older children. This statistical property of viewing could play an important role in integrating multiple views of an object into a unified representation.

Prior to considering the implications of the findings, we replicate the results of Experiment 1 with adults. This replication both validates our head-camera procedure as a way of capturing self-generated views and also provides a measure of the strength of the mature biases given the same objects that children saw and the same measurement procedures.

Experiment 2

Methods

Participants

Participants were 17 adults (7 males and 10 females) with an average age of 23.1 years (range = 20 to 29 years).

Participants were undergraduate or graduate students and were given compensation.

Procedure and design

Participants were told that their task was to learn each object within a small fixed period of time and that they would be tested later in a visual recognition task. The purpose of these instructions was only to encourage attention to the objects; no recognition task was administered. Participants saw the entire stimulus set (16 objects) and were allowed to manipulate each object for 20 seconds. This duration of object exploration is comparable to those in several adult studies (James et al., 2001; Perrett et al., 1992) and to the average inspection times of the children. For calibration, an experimenter placed the head-mounted camera apparatus low on the participant's forehead. For calibration, the experimenter asked the participant to hold a rectangular box and orient one of the faces so that it was completely flat to his line of sight, in the center of his visual field, and aligned to his horizontal axis. A second experimenter watching the head-camera recording gave instructions to adjust the camera so that the viewed object was centered and aligned with the participant's horizontal axis. All other aspects of procedure and data analysis were the same as in Experiment 1.

Results and discussion

Using the same definition of planar and $3/4$ view categories as in Experiment 1, adults showed a strong preference for planar views, with these views constituting 28.7% of the head camera frames which is considerably more than the 5.6% expected given random selection of views, $t(16) = 13.3$, $p < 0.001$. Adults self-presented $3/4$ views only 20.2% of the time, which is reliably less than the expected 33.1% given a random selection of views, $t(16) = -12.2$, $p < 0.001$. Figure 7 shows that the adult dwell time map is highly structured, tightly focused on the planar views of an object held in its upright orientation and with views organized around one axis of rotation. Adults chose upright views 72% of the time, more than then 50% expected by random selection of views, $t(16) = 15.0$, $p < 0.001$. Thus, the biases emerging in children between 12 and 36 months of age strongly govern adults' self generated views of objects. This fact in and of itself emphasizes the importance of understanding the development and role of these biases in object recognition.

General discussion

Perceiver-selected views of objects are biased. The present results show that these biases develop markedly

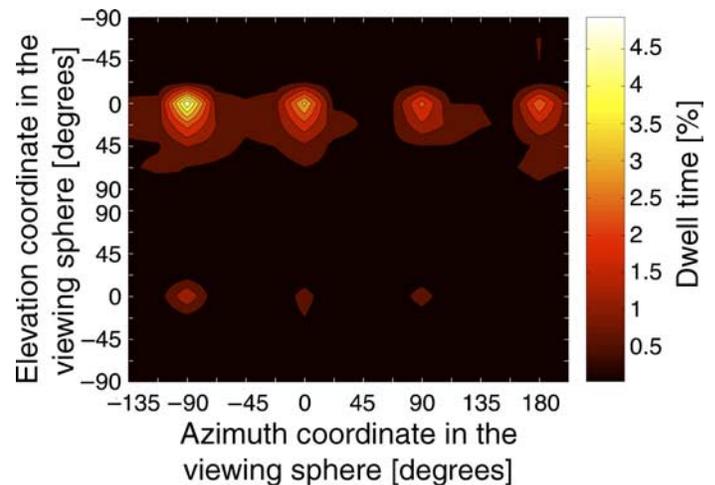


Figure 7. Dwell time map across all objects for the adult group in Experiment 2. The map is shown with the elevation axis duplicated; the top half corresponds to the object in an upright orientation and the bottom half shows the reverse. The order of the elevation values in this axis is arranged so that the object view is symmetrical about the map's horizontal mid-line.

during the period between 12 and 36 months, a significant period of development change on many fronts—in goal directed actions on objects, in object name and category learning, and in visual object recognition. The following discussion considers first, how these developmental changes in biased views may relate to changes in visual object recognition, and subsequently the developmental origins of the biases themselves and their possible relations to other developmental changes during this period.

Explanations of mature visual object recognition fall into two major classes. One class of theories, so-called view-based theories, propose, in a manner consistent with exemplar-based theories of categorization (Palmeri & Gauthier, 2004) that perceivers store viewpoint-dependent surface and/or shape information (Tarr & Vuong, 2002). Under this class of theories, perceivers recognize objects by their similarity to stored views (Bülhoff & Edelman, 1992; Riesenhuber & Poggio, 1999; Tarr & Bülhoff, 1995). Biases in view selection, and developmental changes in those biases (whatever the cause of those changes might be) are clearly relevant to understanding the object recognition system under this class of theories. A second class of theories, so-called object-centered theories, proposes that perceivers represent the geometric structure of the whole object as a spatial arrangement of the major parts (Biederman, 1987; Hummel & Biederman, 1992; Marr, 1982). Although view-based experiences play (at best) a minimal role in this theory, a recent computational model (offered in response to the findings showing developmental changes in object recognition) proposes that the representational elements (geons) of object-centered representations are learned via category learning

(Doumas & Hummel, 2010). Traditionally, view-based and object-based theories were seen as direct competitors (Hummel, 2000; Tarr, Williams, Hayward, & Gauthier, 1998); however, there is a growing consensus that both kinds of representations contribute to human object recognition (Barenholtz & Tarr, 2007; Peterson, 2001; Smith, 2009).

The viewing biases observed here, active viewing itself, and the observed developmental growth of those viewing biases may be the empirical path to unifying the two classes of theories (Savarese & Fei-Fei, 2007; Xu & Kuipers, 2009). For both kinds of theories, the principal axis of an object (typically, the axis of elongation) is central to aligning viewed objects with internal representations. Graf (2006) has also suggested that this alignment process is also the mechanism through which separate views are integrated into unified object-centered representations. The biases observed in children's viewing—the bias for the planar, the bias for viewing objects in a standardized (upright) orientation, and sampling views around one axis or rotation—may all play a critical role in aligning a particular object to those in memory and in integrating those views into a unified representation. The bias for planar views could prove to be particularly important to this integration as these views have little perspective distortion and especially for the planar view that yields the most elongated view, there is reduced variability of information about object shape with small changes in viewing direction (Cutzu & Tarr, 2007; Rosch, Simpson, & Miller, 1976). Thus, planar views may serve stable landmarks in the dynamic experience of the object supporting alignment of different views and their integration (Savarese & Fei-Fei, 2007; Xu & Kuipers, 2009). Sampling views around a single axis of rotation and a bias for a standardized orientation of the object should also particularly benefit aligning multiple views of the same object. In brief, these early viewing biases may play a key role in the development of both view-based and object-based representations. There is already suggestive evidence that the biases benefit adult object recognition (James et al., 2001); a next step is to determine if they do so for young children and whether this benefit depends on the strength of the biases in children's own self-generated object views.

The empirical study of the development of these biases should help illuminate other developmental phenomena in the study of visual object recognition. The extant evidence suggests that the developmental path to mature object recognition is protracted, beginning early in infancy (Johnson, 2001) but not mature until middle childhood or later (Abecassis, Sera, Yonas, & Schwade, 2001; Mash, 2006; Pereira & Smith, 2009; Smith, 2003) with many potential influences (Smith & Pereira, 2009). The path toward unified object-centered representations begins early (Moore & Johnson, 2008; Shuwairi, Albert, & Johnson, 2007; Soska & Johnson, 2008) and initially may be linked to the visual tracking of objects (Johnson,

Amso, & Slemmer, 2003) and later to manual actions on objects (Soska et al., 2010). However, other evidence suggests that object-centered representations of common objects—chairs, cars, horses—emerges later during the same period in which children experience and learn the full range of instances that belong in these common categories (Smith, 2003). Prior to 2 years of age, recognition of instances of common categories may be strongly view-based and dependent on local surface features and piecemeal parts with little dependence on overall shape information (Pereira & Smith, 2009; Quinn, 2004; Rakison & Butterworth, 1998).

The developmental link between the ability to recognize sparse “geon-like” representations (Biederman, 1987) of common objects and early category learning is strong and intriguing. The evidence suggests that children recognize common objects given sparse “geon-like” representations only after they have acquired the names for about 100 common object categories (Pereira & Smith, 2009; Smith, 2003; Son, Smith, & Goldstone, 2008); moreover, recognition of common objects give these minimalist representations of whole object shape is delayed in children who are delayed in early object name learning (Jones & Smith, 2005). The open question is how these changes in visual object recognition related to the biases in the dynamic viewing of objects, biases that are developing in the same period that children shift from object recognition based on piecemeal fragments to object recognition based on overall shape. The temporal correspondence of these two developmental trends suggests a possible developmental dependency: More biased view selection could build more unified whole object representations; alternatively, more structured representations may encourage more structured viewing.

The observed bias for viewing objects in an upright orientation is late emerging and seems likely to depend on experiences with the canonical views and with the functions of objects. The bias for planar views also increases considerably with age and thus seems likely to depend at least in part on experiences with objects as well. However, there is a measurable bias for self-generated planar views even amidst the more disorganized viewing of one-year olds. Given this result, one possibility is that the visual system prefers these views from early infancy. That is, even before infants can hold objects and actually orient them to maximize these properties. By this hypothesis, the development in the strength of these biases might depend on children's developing manual skills as well their experiences with multiple views of objects and object categories. Critically, manipulating objects, in and of itself, introduces constraints and biases object views. This is because real physical objects can be held and manipulated only in some ways and are *efficiently* held and rotated in even fewer ways.

In conclusion, visual object recognition depends on the specific views of experienced objects. These views depend on the perceiver's actions. The nature of early self-

generated object views, and the nature of developmental changes in those views, is critical to understanding how and why the mature object recognition system has the properties that it does. The present results show that there are significant changes in the structure of early self-generated object views with those views becoming more structured and more biased during a developmental period that has been marked as one of significant change in object recognition. Although there are many open questions, these findings point to new empirical directions and new opportunities for advancing understanding of human object recognition.

Acknowledgments

This research was partially supported by National Institute for Child Health and Development (NICHD 28675). AF was partially supported by a Portuguese Ministry of Science and Higher Education PhD scholarship SFRH/BD/13890/2003 and a Fulbright fellowship.

Commercial relationships: none.

Corresponding author: Alfredo F. Pereira.

Email: afpereir@indiana.edu.

Address: Department of Psychological and Brain Sciences, 1101 East Tenth Street Bloomington, IN 47405, USA.

References

- Abecassis, M., Sera, M. D., Yonas, A., & Schwade, J. (2001). What's in a shape? Children represent shape variability differently than adults when naming objects. *Journal of Experimental Child Psychology*, *78*, 213–239.
- Aslin, R. N. (2009). How infants view natural scenes gathered from a head-mounted camera. *Optometry and Vision Science*, *86*, 561–565.
- Barenholtz, E., & Tarr, M. J. (2007). Reconsidering the role of structure in vision. *Psychology of Learning and Motivation*, *47*, 157–180.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–117.
- Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception*, *28*, 575–599.
- Boutsen, L., Lamberts, K., & Verfaillie, K. (1998). Recognition times of different views of 56 depth-rotated objects: A note concerning Verfaillie and Boutsen (1995). *Perception & Psychophysics*, *60*, 900–907.
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, *89*, 60–64.
- Bushnell, E. W., & Boudreau, J. P. (1993). Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. *Child Development*, *64*, 1005–1021.
- Cutzu, F., & Tarr, M. J. (2007). *Representation of three-dimensional object similarity in human vision*. Paper presented at the SPIE Electronic Imaging: Human Vision and Electronic Imaging II, San Jose, CA.
- 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.
- Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, *132*, 920.
- Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*, 1315–1318.
- Harries, M. H., Perrett, D. I., & Lavender, A. (1991). Preferential inspection of views of 3-D model heads. *Perception*, *20*, 669–680.
- Hummel, J. E. (2000). Where view-based theories break down: The role of structure in human shape perception. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 157–185). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.
- 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.
- James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments & Computers*, *34*, 383–390.
- Johnson, S. P. (2001). Visual development in human infants: Binding features, surfaces, and objects. *Visual Cognition*, *8*, 565–578.
- Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 10568–10573.

- 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.
- Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science: A Multidisciplinary Journal*, *32*, 1099–1132.
- Kuipers, J. B. (2002). *Quaternions and rotation sequences: A primer with applications to orbits, aerospace, and virtual reality*. Princeton, NJ: Princeton University Press.
- Landau, B., Hoffman, J. E., & Kurz, N. (2006). Object recognition with severe spatial deficits in Williams syndrome: Sparing and breakdown. *Cognition*, *100*, 483–510.
- Locher, P., Vos, A., Stappers, P. J., & Overbeeke, K. (2000). A system for investigating 3-D form perception. *Acta Psychologica*, *104*, 17–27.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY, USA: Henry Holt and Co., Inc.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences*, *269*–294.
- Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*, 128–152.
- Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066.
- Niemann, T., Lappe, M., & Hoffmann, K. P. (1996). Visual inspection of three-dimensional objects by human observers. *Perception*, *25*, 1027–1042.
- Ornkloo, H., & von Hofsten, C. (2007). Fitting objects into holes: On the development of spatial cognition skills. *Developmental Psychology*, *43*, 404–415.
- Palmer, S., Rosch, E., & Chase, P. (1981). *Canonical perspective and the perception of objects*. Paper presented at the Attention and Performance IX: Proceedings of the Ninth International Symposium on Attention and Performance, Jesus College, Cambridge, England.
- Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*, 291–303.
- Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*, 75–96.
- Pereira, A. F., & Smith, L. B. (2009). Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental Science*, *12*, 67–80.
- Perrett, D. I., & Harries, M. H. (1988). Characteristic views and the visual inspection of simple faceted and smooth objects: “Tetrahedra and potatoes.” *Perception*, *17*, 703–720.
- 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*, 497–497.
- Peterson, M. A. (2001). Object perception. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (pp. 168–203). Oxford, UK: Blackwell.
- Picard, D., & Durand, K. (2005). Are young children’s drawings canonically biased? *Journal of Experimental Child Psychology*, *90*, 48–64.
- Quinn, P. C. (2004). Development of subordinate-level categorization in 3- to 7-month-old infants. *Child Development*, *75*, 886–899.
- Rakison, D. H., & Butterworth, G. E. (1998). Infants’ use of object parts in early categorization. *Developmental Psychology*, *34*, 49–62.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.
- Rock, I. (1973). *Orientation and form*. New York: Academic Press.
- Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 491–502.
- Ruff, H. A., & Saltarelli, L. M. (1993). Exploratory play with objects: Basic cognitive processes and individual differences. In M. H. Bornstein & A. W. O’Reilly (Eds.), *The role of play in the development of thought. New directions for child development, No. 59: The Jossey-Bass education series* (pp. 5–16). San Francisco, CA, US: Jossey-Bass.
- Ruff, H. A., Saltarelli, L. M., Capozzoli, M., & Dubiner, K. (1992). The differentiation of activity in infants’ exploration of objects. *Developmental Psychology*, *28*, 851–861.
- Savarese, S., & Fei-Fei, L. (2007). *3D generic object categorization, localization and pose estimation*. Paper presented at the IEEE 11th International Conference on Computer Vision.
- Shen, H., Baker, T., Candy, R., Yu, C., & Smith, L. B. (in preparation). Head and eye stabilization during reaching by toddlers.
- Shuwairi, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. *Psychological Science*, *18*, 303–307.
- Smith, L. B. (1999). Children’s noun learning: How general learning processes make specialized learning mechanisms. *Emergence of Language*, *277*–303.

- Smith, L. B. (2003). Learning to recognize objects. *Psychological Science, 14*, 244–250.
- Smith, L. B. (2009). From fragments to geometric shape. *Current Directions in Psychological Science, 18*, 290–294.
- Smith, L. B., & Pereira, A. F. (2009). Shape, action, symbolic play, and words: Overlapping loops of cause and consequence in developmental process. In S. P. Johnson (Ed.), *A neo-constructivist approach to early development*. New York: Oxford University Press.
- Smith, L. B., Yu, C., & Pereira, A. F. (2010). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, in press.
- 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 3D object completion. *Developmental Psychology, 46*, 129–138.
- Soska, K. C., & Johnson, S. P. (2008). Development of three-dimensional object completion in infancy. *Child Development, 79*, 1230.
- Street, S. Y., James, K. H., Jones, S. S., & Smith, L. B. (in press). Toward an understanding of coordinated dorsal and ventral stream functions: Vision for action in toddlers. *Child Development*.
- Tarr, M. J., & Bülthoff, 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*, 1494–1505.
- Tarr, M. J., & Vuong, Q. C. (2002). Visual object recognition. In H. Pashler & S. Yantis (Eds.), *Steven's handbook of experimental psychology (3rd ed.)*, Vol. 1: *Sensation and perception* (pp. 287–314). Hoboken, NJ, US: John Wiley & Sons, Inc.
- Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. *Nature Neuroscience, 1*, 275–277.
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11*, 58–64.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use on classification. *Nature Neuroscience, 5*, 682–687.
- Xu, C., & Kuipers, B. (2009). *Construction of the object semantic hierarchy*. Paper presented at the Fifth International Cognitive Vision Workshop (ICVW 2009).
- Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy, 13*, 229–248.