Developmental Science 7:5 (2004), pp 543–549

REPORT

Two- and three-year-olds infer and reason about design intentions in order to categorize broken objects

Deborah G. Kemler Nelson, Morghan B. Holt and Louisa Chan Egan

Department of Psychology, Swarthmore College, USA

Abstract

In naming artifacts, do young children infer and reason about the intended functions of the objects? Participants between the ages of 2 and 4 years were shown two kinds of objects derived from familiar categories. One kind was damaged so as to undermine its usual function. The other kind was also dysfunctional, but made so by adding features that appeared to be intentional. Evidence that 2-, 3- and 4-year-olds were more likely to apprehend the broken objects than the intentionally dysfunctional objects as members of the familiar lexical categories favors the conclusion that, in naming, children may spontaneously infer and reason about design intentions from an early age. This is the first evidence that 2- and 3-year-olds not only take design intentions into account in object categorization, but that they do so even without explicit mention of the objects' accidental or intentional histories. The results cast doubt on a proposal that young children's lexical categorization is based on automatic, non-deliberative processes.

Introduction

The current study is motivated by two recent lines of work concerning young children's naming of artifacts – lines of work that are related to each other in an interesting way. One has to do with the role of functional information in children's artifact labeling. This has been a very controversial area, replete with apparently conflicting experimental results and conclusions concerning the dominance of appearances (or shape, in particular) versus functional information. For example, Smith, Jones and Landau (1996) and Landau, Smith and Jones (1998) demonstrated that children younger than 5 years of age generalized the newly learned name of a novel artifact to objects more like it in appearance than function. They concluded that young children's lexical categorization is largely immune to functional information, a claim central to their 'dumb attentional mechanism' (DAM) account. Contrasting results and conclusions were offered by Kemler Nelson and her colleagues. In the same kind of experimental comparison, pitting appearance against functional similarity, these investigators found that preschoolers (Kemler Nelson & 11 Swarthmore College students, 1995; Kemler Nelson, Frankenfield, Morris & Blair, 2000a) and even toddlers (Kemler Nelson, 1999; Kemler Nelson, Russell, Duke & Jones, 2000b) weighed functional information over appearance similarity. Accordingly, they argued that functional information is as important to young children's artifact naming as it is to adults'.

The other line of work relevant to the current study concerns children's tendency to take into account the intentions of artifact designers in lexical categorization. For example, Gelman & Bloom (2000) showed 3-year-olds variants of simple, common objects that were described as having been created intentionally or by accident. When the children were asked, 'What is this?', they were more likely to provide artifact names for the objects described as intentionally designed than those described as having come to exist accidentally. Furthermore, Diesendruck, Markson and Bloom (2003) showed that generalization of new names by shape similarity was undermined by describing the demonstrated functions of novel artifacts as intended by the object designers.

How might these two lines of work be connected? We have suggested that a satisfying resolution of the controversy concerning the importance of appearance versus functional information in artifact categorization can be reached by considering that when children are shown new functions for new artifacts, they privilege functional information (over appearance) specifically when it provides

Address for correspondence: Deborah G. Kemler Nelson, Department of Psychology, Swarthmore College, Swarthmore, PA 19081-1390, USA; e-mail: dkemler1@swarthmore.edu

a compelling account of the intention of the object designer. Kemler Nelson *et al.* (2000a) experimentally manipulated whether the novel functions demonstrated for artifacts were plausible or implausible accounts of the origins of the objects. Four-year-olds generalized names more often for functionally alike, but perceptually less similar objects only when the functions that were demonstrated for the original objects were judged by adults to be plausible accounts of what the objects had been designed for. Children were more likely to dismiss the functional information when it could not underwrite a coherent causal account of the objects' origins. Unwittingly, investigators such as Smith *et al.* (1996), who have provided evidence against the importance of function, may have created materials of the latter sort. (For related proposals see Bloom, 2000 and Gelman & Bloom, 2002.)

Recently, Kemler Nelson *et al.* (2002) offered further evidence that spontaneous inferences about design intentions mediate children's lexical categorization. The focus was the categorization of broken objects – interesting because damage may undermine objects' capacities to function in intended ways. If categorizers prioritize intended function over current status, they should accept a damaged object as a category member – such as, believing a chair with a broken leg is still a chair (Bloom, 1996). Kemler Nelson *et al.* (Experiment 4) created objects by modifying exemplars of familiar categories. Two dysfunctional instances of each category were constructed: one, an object damaged so as to undermine the function for which it was designed; the other, an object with an added feature that appeared to originate in intentional design, but that undermined the usual function. Even 4-year-olds were sensitive to design intentions when they named: more broken objects were labeled as category members than intentionally dysfunctional objects.

Do still younger children consider design intentions to be relevant to categorization? Although previous studies have suggested they do (Diesendruck *et al.*, 2003; Gelman & Bloom, 2000), children in these studies were provided with explicit information about intentions rather than requiring them to use inferential processes. Accordingly, the demonstrations are limited in two ways. First, the conditions were unusual in that explicit information about the intentional or accidental origins of the objects was provided – information rarely available in natural contexts. Second, although the children had to reason from information about the objects' histories, they did not have to infer these histories by themselves. The kind of comparison arranged by Kemler Nelson *et al.* (2002) is, by comparison, a particularly challenging test for the role of intended function in artifact categorization.

Until now, such a test had never been arranged for children under 4 years old and, accordingly, the evidence for inferential processes in their categorization is limited to demonstrations that they consider current functional status (Kemler Nelson, 1999; Kemler Nelson *et al.*, 2000b). Yet, this is exactly the age group for whom the DAM account, offered by Smith *et al.* (1996), seems most compelling. Reviewing relevant findings, Landau *et al.* (1998) write, '(T)he results suggest that immunity to functional knowledge should be seen earlier than age 3, but may begin to change with development after this point' (p. 5). Evidence that children 3 years old and younger not only use functional information, but use information about *intended* function, would seem to constitute dramatic counter-evidence to this conclusion.

In the present study, we build on the methods of Kemler Nelson *et al.* (2002) to investigate the categorization of broken artifacts by toddlers and young preschoolers. If these children spontaneously infer and weigh information about design intentions when they categorize objects, then – like older children and adults – they should be more likely to include damaged than intentionally dysfunctional objects as category members. Such a finding would add credibility to arguments that, even from an early age, lexical categorization may involve significant amounts of deliberative and inferential processing.

In the earlier study, children were simply presented with objects one-by-one and asked what they would call them. Here, we supplemented this naming test with a forced-choice test: children chose between the accidentally and intentionally dysfunctional objects of each category the one that was a category member. The order of the tests was counterbalanced for 2- and 3-year-old children. For comparison, we also included a group of 4-year-olds who got the naming test first.

Method

Participants

Participants were 32 2-year-olds ($M = 2$ years, 7 months with a range of 2 years, 0 months to 3 years, 0 months), 32 3-year-olds ($M = 3$ years, 7 months with a range of 3 years, 1 month to 3 years, 11 months), and 16 4-year-olds ($M = 4$ years, 5 months with a range of 4 years, 2 months to 4 years, 11 months). There were approximately equal numbers of boys and girls, drawn from a suburban upper-middle-class community.

Stimulus materials

Eight pairs of test objects derived from familiar categories (Kemler Nelson *et al.*, 2002) were used: combs, cups,

Table 1 *Description of the properties modified to undermine function and perceptual similarity ratings of the resulting objects*

Type of object	Accidentally dysfunctional	Intentionally dysfunctional
Combs	3.58 Most teeth partially or fully broken off	3.58 Edges sealed by transparent band of plastic
Cups	5.50 Irregular piece broken out from rim to base	5.25 Circular hole, rimmed in metal, centered in the base
Flip-flop sandals	3.58 Thong pulled out from between-toe position, and chunks removed from sole	3.67 Two small wooden balls attached to sole
Forks	3.58 Tines broken off or bent irregularly	4.42 Shaft and tines totally flattened
Hats	4.33 Rips running all the way from top center to brim	5.67 Crown, stuffed with cotton, and sewn shut with same fabric as hat
Safety pins	3.58 Receiving arm bent and other arm partially broken off	4.42 Constructed from a single, continuous piece of wire
Scissors	3.50 One finger hole broken off and one blade partially broken off	6.00 Finger holes glued permanently together, and only a single blade
Shovels	3.50 Large piece of scoop broken out	4.75 Shaft replaced by flexible plastic tubing

Note: Ratings were obtained by Kemler Nelson *et al.* (2002). Higher numbers indicate greater similarity.

flip-flop sandals, forks, hats, safety pins, scissors and shovels. One member was damaged so it could not function as it was originally designed to do. Objects of this sort appeared to lack the ability to function as members of their categories for accidental reasons. The other member was also modified by structural changes that undermined function, but these modifications appeared to derive from intentional design. These objects invited the inference that they were not intentionally designed to function in a way that was typical of the category whose members they most resembled. For example, the damaged cup was created by breaking off an irregular piece of a plastic cup such that it could not hold liquid. The intentionally dysfunctional cup was created by drilling a circular hole, centered in the cup bottom, and edging the hole with a metal ring. Thus, both members of the pair were dysfunctional as cups, but one appeared to be accidentally dysfunctional and the other intentionally dysfunctional. A full description of the modifications that undermined function in the two different ways is provided in Table 1.

In creating the test objects, we tried to make the broken object in each pair no more similar in appearance to members of its original category than its intentionally dysfunctional counterpart. Accordingly, the damaged objects were occasionally derived from less typical category members than the intentionally dysfunctional objects (for the cups, combs and forks). In addition, the modifications of the intentionally dysfunctional objects were as minimal as possible, while undermining the relevant function, but, in two cases, additional features of these objects were added to hint at a different function (loops attached to the hat; a handle-shaped garden spray attachment to the shovel).

For purposes of the test arranged here, it was important that the accidentally dysfunctional objects were no more perceptually typical of the categories from which they were derived than the intentionally dysfunctional objects. Accordingly, Kemler Nelson *et al.* (2002) obtained ratings from 12 college students. Students used a 7-point scale, ranging from 1 (hardly at all) to 7 (almost exactly), to rate the extent to which the appearance (and only the appearance) of the object was similar to that of its typical, everyday counterpart (e.g. 'a typical cup'). They were explicitly asked to make their ratings without regard to functional considerations. Students were asked to pretend that they were aliens who had learned about Earth only through pictures of Earth artifacts. 'You've seen 3-D images of various objects, but you have no idea what they do or how they work,' and, in addition, 'Some of the objects are broken, but I want you to evaluate their appearance as it is now, not as what you imagine it used to be.' The mean ratings for each object are shown in Table 1.

As previously reported by Kemler Nelson *et al.* (2002), an analysis of variance with category and type of object (accidentally or intentionally dysfunctional) revealed that the intentionally dysfunctional objects ($M = 4.73$) were actually rated as more similar in appearance to members of their category than the accidentally dysfunctional objects ($M = 3.92$), $F(1, 11) = 15.61$, $p < .005$, although a reliable interaction of category with type of object, $F(7, 77) = 4.83$, $p < .025$, suggested that the difference was

not uniform across categories. Of critical importance, in no case was there evidence that appearance similarity favored the broken object in a pair. Only for the cups did the mean rating for the broken object exceed that for the intentionally dysfunctional object, but raters were inconsistent: five showed this pattern, three showed the reverse and four rated the two objects identically.

Two additional object pairs were used as control pairs in the test phase: a book and a damaged block, and eyeglasses and a novel object that was damaged. They were added to ensure that a response bias toward broken objects could not account for the critical patterns of results. In the forced-choice phase, children were asked which object in each of these control pairs was a 'book' and 'a pair of eyeglasses', respectively.

Six objects used in a warm-up phase were a torn envelope, a pencil missing an eraser, a 'crazy' straw and three novel items created for previous studies. Two of these unfamiliar objects appeared to be damaged. The warm-up objects were designed to set up various expectations: that some objects might be damaged, damage might or might not undermine function, some objects might be atypical and yet nameable by a common category name, and some might not be nameable in a conventional way at all. Importantly, children were told that it was acceptable to reply 'I don't know' when shown an object for which they did not have a ready name.

Design

All sessions began with a phase in which participants were asked to name the six warm-up objects, presented in random order. The naming test and then the forced-choice test followed for half the 2- and 3-year-olds and all the 4-year-olds. The remaining participants received the forced-choice test before the naming test. For the naming test, the 16 test objects were divided into two complementary sub-blocks, each consisting of the damaged objects from half the categories and the intentionally dysfunctional objects from the other half. The broken object from one control pair and the intact object from the other control pair were also assigned to each sub-block. The assignment of individual items to sub-blocks and the order of objects within sub-blocks were determined at random for each participant. In the forced-choice test, the order of object pairs and the positions of the accidentally dysfunctional objects were also randomly determined.

Procedure

At the beginning, participants were told they would be seeing objects, some familiar and some new. They were instructed to tell the experimenter 'what you would call' each thing as it was handed to them for inspection. They were informed that they might not know a name for every object, and it was acceptable to say 'I don't know' when they did not know the name. On the few occasions that a child gave an ambiguous response (e.g. 'Something that used to be an envelope'), the experimenter followed up with a further probe (e.g. 'Is it still an envelope now?'). The same procedure was used in the naming test.

On each forced-choice trial, the two objects of each pair were presented side-by-side, and the child heard, for example: 'One of these is a cup and one of these is not a cup. Can you show me the one that is a cup?' The category was named by the experimenter. When the forced-choice test preceded the naming test, each trial began by successively handing the two objects to the child for inspection in random order before placing them side-by-side for choice. However, when the forced-choice test was second, the two objects which the child had inspected earlier were simply placed side-by-side.

Results

Coding of the naming responses

Naming responses were coded into three categories: (1) 'inclusions' – names implying category membership, whether or not modified by a description that implied damage (e.g. 'a chewed up shovel' or 'a cup that can't hold water'); (2) 'non-inclusions' – responses explicitly labeling an alternative category, explicitly denying membership in the relevant category, referring to the object's material, or 'I don't know'; and (3) 'ambiguous' (e.g. 'something that looks like a hat'). An 'I don't know' response was coded as a 'non-inclusion' because there was reason to believe the child knew the name of the category from which the object derived. In fact, in a number of cases, a child responded 'I don't know' to one member of a pair, while successfully naming the other member.

When a participant offered multiple responses for an object, a non-ambiguous response was counted over an ambiguous one. A conservative criterion was applied on the rare occasion that both an inclusion and a non-inclusion were offered: a non-inclusion was counted for damaged objects and an inclusion was counted for intentionally dysfunctional objects.

Responses in the naming test

Each category inclusion was counted as 1, each non-inclusion as 0, and the rare ambiguous response as

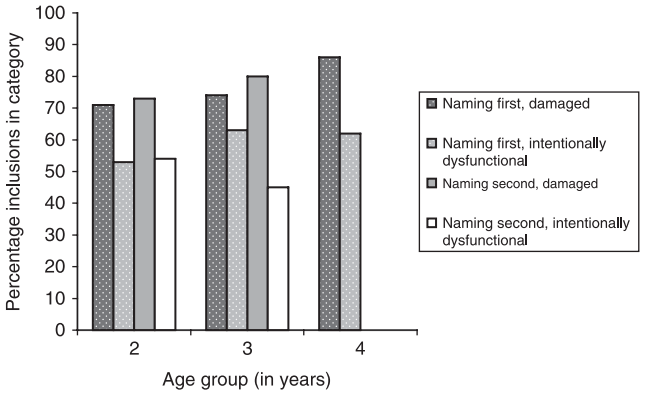

Figure 1 *The percentage of broken and intentionally dysfunctional objects named as members of the category.*

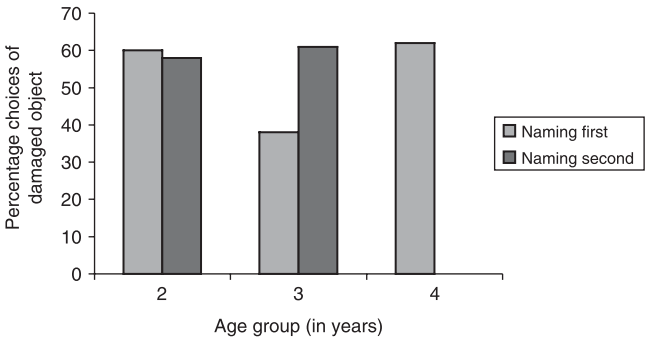

Figure 2 *The percentage of broken objects selected as members of the named category.*

1/2. To calculate the percentage of inclusions for accidentally and intentionally dysfunctional objects, these scores were added and divided by the possible maximum of 8. These percentages are shown in Figure 1.

An analysis of variance (ANOVA) for a 2 (age) × 2 (test order) × 2 (object type) mixed design was carried out on the scores of 2- and 3-year-old children. The analysis revealed a main effect of object type, $F(1, 60) = 70.12$, $p < .001$, an interaction between object type and test order, $F(1, 60) = 5.08$, $p < .05$, and an interaction between object type, test order and age, $F(1, 60) = 5.84$, $p < .025$. The main effect indicates fewer inclusions of the intentionally dysfunctional than the damaged objects. The three-way interaction stems from the tendency of 3-year-olds, but not 2-year-olds, to provide fewer names of intentionally dysfunctional objects when naming followed forced-choice than when naming was first. Apparently, being told in the forced-choice test that only one of the two paired objects was a category member affected how the 3-year-olds named. An additional ANOVA included the 4-year-old data and the scores of the younger children with the naming test first. It yielded a main effect of object type, $F(1, 45) = 62.33$, $p < .001$, due to more inclusions of damaged than intentionally dysfunctional objects, and an interaction between object type and age, $F(2, 45) = 5.73$, $p < .01$, indicating that the difference due to object type was not uniform across the three ages.

The interactions notwithstanding, the most important findings from the naming test can be discerned by inspecting each group independently. For each combination of age and test order, an independent *t*-test revealed that more damaged than intentionally dysfunctional objects were named as category members, each $t(15) \geq 2.75$, $p < .02$. Of the 80 children, only three showed numerically more inclusions of intentionally dysfunctional than accidentally dysfunctional objects, and 65 favored the broken objects. Importantly, even when object type was the random variable in independent *t*-tests, the same pattern of including more damaged than intentionally dysfunctional objects obtained, $t(7) \geq 2.38$, $p < .05$, in each of the five cases. Indeed, for each combination of age and test order, at least seven of the eight categories showed numerically more inclusions of the damaged variant. Hence, the major effect of interest is not being selectively driven by just a few categories of objects. The finding that young children are inferring and using information about intentional design when they name artifacts is general from the age of 2 years upward.

Responses in the forced-choice test

To compute the percentage of forced choices of the damaged objects, choice of the damaged object was counted as 1, choice of the intentionally dysfunctional object as 0, and the infrequent failure to choose as 1/2. The percentages are shown in Figure 2.

An ANOVA for a 2 (age) × 2 (test order) design was applied to the scores of the 2- and 3-year-olds. The analysis yielded an effect of age, $F(1, 60) = 6.43$, $p < .025$, an effect of test order, $F(1, 60) = 8.13$, $p < .01$, and an interaction between the two variables, $F(1, 60) = 10.72$, $p < .005$. An ANOVA using only the responses from children who received the forced-choice test second yielded an effect of age, $F(2, 45) = 12.24$, $p < .001$. Again, the most important results are found by looking at each of the groups independently. With the exception of the 3-year-olds with the forced-choice test second, each group showed a reliable tendency to select the broken objects as the category members, each $t(15) > 2.27$, $p < .05$. The exceptional group of 3-year-olds actually showed an

opposite tendency, $t(15) = 2.35$, $p < .05$.[1] Hence, in four of five cases, the implications of the forced-choice test converge with the wholly consistent findings of the naming test: children 2 to 4 years old are more likely to label damaged than intentionally dysfunctional objects as category members.

Discussion

The present findings constitute evidence that 2- and 3-year-old children not only take functions into account when they name objects, but that they also take *intended* functions into account. Notably also, children's consideration of design origins here was entirely spontaneous. Previous demonstrations that matters of intentional design have an impact on young children's categorization of objects have involved manipulations in which children were informed about objects' intentional or accidental histories (Diesendruck *et al.*, 2003; Gelman & Bloom, 2000). In the current study, in the absence of such an explicitly imposed distinction, children were inclined on their own to infer these histories from the existing structures of the objects, and to weigh these inferences in their categorization. Although it is unsurprising that young children could detect damage and understand it to be accidental, it is notable that they discounted this damage when categorizing, even though it undermined current function. Still more remarkable is that, for objects whose overall perceptual features were at least as typical of category members as their damaged counterparts, young children sometimes inferred that their dysfunctional features were intentional, and, consequently, sufficient to undermine category membership.

What should one make of the finding that participants, while generally favoring the inclusion of the damaged objects, often also named the intentionally dysfunctional objects with the category label? On one interpretation, this is a theoretically uninteresting consequence of their not having any other ready name for the object with which to address the experimenter's request. Generally, children seemed somewhat hesitant to respond 'I don't know.' However, on a different interpretation, the fact that the intentionally dysfunctional objects had so many perceptual features typical of the category from which they were derived may have counted as a factor in children's considerations of intentional design, in addition to considerations of intended function (Bloom, 1996). Which interpretation, if either, is the correct one cannot be decided on the basis of the current evidence.

In combination, the contrasting patterns of naming accidentally and intentionally dysfunctional artifacts support a conclusion that the process of lexical categorization of artifacts may be deliberative even in children as young as 2 years of age (see also Booth & Waxman, 2002; Gopnik & Sobel, 2000). Were the children in the present study simply naming on the basis of strong, automatic pulls from perceptual features and well-learned associations (as postulated in the DAM account of Smith *et al.*, 1996), it would be particularly surprising that they sometimes refrained from naming the intentionally dysfunctional objects with the labels of the familiar categories whose members they most closely resembled. It is reasonable to hypothesize that children would have to inhibit such automatic control of attention in order to withhold the familiar labels in these cases. The fact that the children did not always do so may indicate that these inhibitory processes were not always strong enough to block the familiar name. Still, the more important finding – namely, that the children attached the familiar name differentially to the intentionally and accidentally dysfunctional objects – provides some of the strongest evidence to date that toddlers and young preschoolers sometimes name in a deliberative and non-automatic way.

[1] It is difficult to come up with a convincing explanation of the anomalous result that the 3-year-olds with the forced-choice test second responded differently in that test from (a) the way they responded in the naming test, (b) the 3-year-olds who got the forced-choice test first, and (c) both the 2-year-olds and the 4-year-olds who got the forced-choice test second. One might suggest that children at this age (but only this age) took the administration of the second test as an indication that their response tendency in the first test was in need of modification, but this is pure speculation.

Acknowledgements

Support for this work was provided by Swarthmore College in the form of Faculty Research Grants and a sabbatical leave award to Deborah G. Kemler Nelson, and Joel Dean Summer Research Assistantships to the junior authors. We thank the parents, staff and children who participated in the study, including those from Trinity Cooperative Day Care Center, Swarthmore Presbyterian Day Nursery School and Kehillah Israel Summer Day Camp. We appreciate the assistance of Don Reynolds in modifying the stimulus objects.

References

Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, **60**, 1–29.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Booth, A.E., & Waxman, S.R. (2002). Word learning is 'smart': evidence that conceptual information affects preschoolers' extensions of novel words. *Cognition*, **84**, B11–B22.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on intentions in naming artifacts. *Psychological Science*, **14**, 164–168.

Gelman, S., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding how to name it. *Cognition*, **76**, 91–103.

Gopnik, A., & Sobel, D. (2000). Detecting blickets: how young children use information about novel causal powers in categorization and induction. *Child Development*, **71**, 1205–1222.

Kemler Nelson, D.G. (1999). Attention to functional properties in toddlers' naming and problem-solving. *Cognitive Development*, **14**, 77–100.

Kemler Nelson, D.G., & 11 Swarthmore College students (1995). Principle-based inferences in young children's categorization: revisiting the impact of function on the naming of artifacts. *Cognitive Development*, **10**, 347–380.

Kemler Nelson, D.G., Frankenfield, A., Morris, C., & Blair, E. (2000a). Young children's use of functional information to categorize artifacts: three factors that matter. *Cognition*, **77**, 133–168.

Kemler Nelson, D.G., Russell, R., Duke, N., & Jones, K. (2000b). Two-year-olds will name artifacts by their functions. *Child Development*, **71**, 1271–1288.

Kemler Nelson, D.G., Herron, L., & Morris, C. (2002). How children and adults name broken objects: inferences and reasoning about design intentions in the categorization of artifacts. *Journal of Cognition and Development*, **3**, 301–332.

Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, **38**, 1–27.

Smith, L.B., Jones, S.S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, **60**, 143–171.

Received: 23 May 2003
Accepted: 14 January 2004