Child Development, May/June 2003, Volume 74, Number 3, Pages 822–833

Is a Picture Worth a Thousand Words? Preference for Auditory Modality in Young Children

Vladimir M. Sloutsky and Amanda C. Napolitano

Linguistic labels play an important role in young children's conceptual organization: When 2 entities share a label, people expect these entities to share many other properties. Two classes of explanations of the importance of labels seem plausible: a language-specific and a general auditory explanation. The general auditory explanation argues that the importance of labels stems from a privileged processing status of auditory input (as compared with visual input) for young children. This hypothesis was tested and supported in 4 experiments. When auditory and visual stimuli were presented separately, 4-year-olds were likely to process both kinds of stimuli, whereas when auditory and visual stimuli were presented simultaneously, 4-year-olds were more likely to process auditory stimuli than visual stimuli.

Linguistic labels play an important role in young children's conceptual organization and thinking: When two entities share a label, these entities are expected to share many other properties. For example, if a child learned that an animal named *lion* has a certain biological property, the child would expect another lion to have the same biological property.

In previous experiments with young children, auditorily presented linguistic labels were found to support categorization (Balaban & Waxman, 1997; Sloutsky & Fisher, 2001), inductive inference (Gelman & Markman, 1986; Sloutsky, Lo, & Fisher, 2001), and similarity judgment (Sloutsky & Lo, 1999). Suppose there is a triad of stimuli (a target and two test stimuli) in which one of the test stimuli has the same label as the target. If young children are asked either (a) to induce properties from one of the test stimuli to the Target, (b) to group one of the test stimuli with the target, or (3) to judge similarity between each of the test stimuli and the target, then matched linguistic labels make sizable contributions to young children's choices. Furthermore, contributions of matched linguistic labels are often larger than contributions of similar appearances (see Gelman & Markman, 1986; Sloutsky & Lo, 1999; Sloutsky et al., 2001, for related discussions). In other tasks, linguistic labels were found to support categorization by attracting infants' and young children's attention to correspondences among presented objects (Balaban & Waxman, 1997; Markman, 1989; Waxman & Markow, 1995).

Although the prominent role of auditorily presented linguistic labels has been established, mechanisms underlying these effects remain unknown. Two classes of explanations seem plausible: The prominence of labels could stem from language-specific factors, or it could stem from general auditory factors.

The language-specific explanation has two variants: semantic and prosodic. The semantic explanation is predicated by the view that young children assume that entities are members of categories and that labels presented as count nouns convey category membership (Gelman & Coley, 1991). These assumptions lead young children to infer that entities that are denoted by the same count noun belong to the same category (Gelman & Markman, 1986; Markman, 1989; see also Waxman & Markow, 1995, for a discussion). Therefore, according to the semantic explanation, if entities share a label presented as a count noun, this shared count noun suggests that entities belong to the same category (thus supporting categorization). Furthermore, belonging to the same category indicates that members of the category share nonobvious properties (thus supporting inductive inference about these properties).

However, it has been demonstrated that even when labels were digitally edited so that they could not be identified as count nouns, they exhibited

---

Vladimir M. Sloutsky and Amanda C. Napolitano, Center for Cognitive Science, The Ohio State University. This research has been supported by grants from the National Science Foundation (BCS 0078945) to Vladimir M. Sloutsky. We would like to thank George Hollich, Chris Robinson, and Aaron Yarlas for their feedback.

Correspondence concerning this article should be addressed to Vladimir M. Sloutsky, Center for Cognitive Science, 208C Ohio Stadium East, 1961 Tuttle Park Place, The Ohio State University, Columbus, OH 43210. Electronic mail may be sent to Sloutsky.1@osu.edu.

0009-3920/2003/7403-0011

facilitative effects on infants' categorization similar to those of count nouns (Balaban & Waxman, 1997). Therefore, it is possible that facilitative effects of labels might not be limited to semantic effects but that additional effects might be due to infants' and young children's special attention to the prosodic components of human speech that distinguish speech from other sounds. As a result of this special attention, any linguistic labels accompanying entities may attract attention to other properties correlating with the label.

Although both semantic and prosodic effects are plausible, there is also evidence that nonspeech sounds may have effects similar to those of speech sounds. For example, Roberts and Jacob (1991) presented 15-month-old infants with pictures of members of natural kind categories (e.g., a particular kind of animal), such that either a linguistic label or a piece of instrumental music accompanied each picture. When infants habituated, researchers presented them with members of either same or different categories. Results indicated that the instrumental music had the same facilitating effects on categorization for the 15-month-olds as did the linguistic labels. In addition, it has been demonstrated that 13-month-old infants used both words and nonspeech sounds as indexes of objects, although the tendency to use nonspeech sounds weakened by 20 months of age (Woodward & Hoyne, 1999). Therefore, it is possible that the importance of labels stems from general auditory effects: Early in development auditory input may have a privileged processing status (as compared with the status of visual input), and this privileged processing status may extend into early childhood.

There is evidence that in infancy the auditory modality does indeed dominate the visual modality (see Lewkowicz, 1994, for a review). In particular, 6- to 10-month-olds are more likely to process changes in auditorily presented stimuli than in visually presented stimuli (Lewkowicz, 1988a, 1988b, 1994). In these studies, infants were first habituated to compound stimuli (e.g., a flashing checkerboard and a pulsing sound). They were then presented with a series of test trials, some of which had the old visual component and a changed auditory component, some had the old auditory component and a changed visual component, and some had both components changed. Infants readily detected changes in the auditory component but not in the visual component. At the same time, when habituation stimuli consisted only of visual components, even 6-month-olds readily detected change in the visual component (Lewkowicz, 1994), indicating that the previous findings were not simply due to infants' lack of ability to detect changes in visual components.

Although studies with infants present corroborative evidence, there is little research directly examining the possibility that for young children auditory input has a privileged processing status. The goal of research presented here was to fill this gap by directly testing the hypothesis that young children exhibit dominance in the auditory modality.

In what follows we present several experiments designed to test this hypothesis. If the auditory modality has a privileged processing status, whenever visual stimuli are paired with auditory stimuli young children should be more likely to attend to auditory stimuli than to visual stimuli. At the same time, visual stimuli should be fully attended when young children are presented without accompanying auditory stimuli. To achieve this goal, we deemed it necessary to use visual and auditory stimuli that were comparable in their familiarity and discriminablity. Because auditory stimuli were completely novel and it is difficult to label three-sound patterns, we selected visual stimuli that were novel and difficult to label as well (we used novel visual scenes or arrangements of geometric objects). Equal discriminability of auditory and visual stimuli was ascertained in a set of calibration experiments.

In Experiments 1 and 3, we used a modification of the switch design (see Werker, Lloyd, Cohen, Casasola, & Stager, 1998, for another example of this design), in which participants were presented with two stimulus sets, each consisting of an auditory and a visual component, and were trained to select consistently one set over the other. When training was accomplished, participants were presented with a choice of two test sets. The choice sets were created in the following manner. The trained set was split, and the visual component of the trained set was paired with a novel auditory stimulus whereas the trained auditory component was paired with a novel visual stimulus. Unlike in the original switch design (Werker et al., 1998), where only the pairing of visual and auditory stimuli changed, in the current studies the trained visual component was paired with a completely new auditory component and the trained auditory component was paired with a completely new visual component. We argue that if the visual stimulus has a larger attentional weight, participants should select the set pairing the trained visual stimulus with the novel auditory stimulus, whereas if the auditory stimulus has a larger weight, they should select the set pairing the trained auditory stimulus with the novel visual stimulus.

In Experiments 2 and 4, we examined whether young children encode both auditory and visual components. Participants were presented with an immediate recognition task, in which a compound auditory–visual set appeared on screen, followed by a test item. Test items were either identical to the target, had one of the components changed (i.e., auditory or visual), or had both components changed. Participants were asked to make a same–different recognition judgment. It was expected that if participants encode both auditory and visual components, they should accurately accept old targets and accurately reject other items. At the same time, if participants fail to encode either visual or auditory components, they should erroneously accept sets with the new auditory component or new visual component, respectively.

Experiment 1

The goal of Experiment 1 was to determine the hypothesized dominance of the auditory modality in young children.

Method

Participants

Participants were 15 young children ($M = 4.46$ years, $SD = 0.235$ years; 7 boys and 8 girls) recruited from several daycare centers located in middle-class suburbs of Columbus, Ohio, and 20 college undergraduates ($M = 18.45$ years; $SD = 0.887$ years; 4 men and 16 women) from a large midwestern university participating in the experiment in partial fulfillment of a course requirement. The majority of participants were Caucasian.

Materials

Materials consisted of 24 stimulus sets, each composed of a visual and an auditory stimulus. The visual stimuli, which were presented on a computer screen, were 10 cm × 10 cm digitized photographic landscape images, each consisting primarily of a different type of green foliage. These visual stimuli were selected to minimize the familiarity of visual stimuli (another consideration was to eliminate the possibility that participants would be able easily to label them), thus making them comparable to the auditory stimuli. The auditory stimuli were computer-generated patterns, each consisting of three unique simple tones. Simple tones varied on timbre (sine, triangle, or sawtooth) and frequency (between 1 Hz and 100 Hz). Each simple tone was 0.3 s in duration and was separated by .05 s of silence, with total pattern duration of 1 s. The average sound level of auditory stimuli was 67.8 dB (with a range from 66 dB to 72 dB), which is comparable to the sound level of human voice in a regular conversation.

A calibration experiment was performed to ensure that for each modality, all auditory and visual stimuli were sufficiently discriminable. This was accomplished using two same–different immediate recognition tasks (one for auditory stimuli and one for visual stimuli), in which a different sample of 20 undergraduates and 14 children 4 years of age made same–different judgments after being presented with pairs of either auditory or visual stimuli. Within each pair, stimuli were presented successively for 1 s each. The calibration experiment was performed using Superlab Pro 2.0 software (Cedrus Corporation, 1999), which allowed recording of responses and their latency. Participants exhibited near ceiling performance, with adults discriminating auditory stimuli in 96% of trials and visual stimuli in 97% of trials, both above chance, ts$(19) > 44$, ps$ < .0001$, and child participants discriminating auditory stimuli in 86% of trials and visual stimuli in 93% of trials, both above chance, ts$(12) > 15$, ps$ < .0001$.

After ensuring their discriminability, the individual auditory and visual stimuli were configured randomly into paired sets. Each set was composed of a simultaneous visual and auditory component (i.e., each image's presentation lasted for the duration of its accompanying sound), which was created by pairing an auditory stimulus and a visual stimulus so that they were perceived as one unit. The sound and image components were randomly selected from the pool of each stimulus group. There were four types of stimulus sets created: (a) VIS_1AUD_1, the training target set that participants were trained to select; (b) VIS_2AUD_2, which was presented as a distracter across the training trials; (c) VIS_1AUD_{new}, which matched the training target's visual component but had a novel auditory component; and (d) $VIS_{new}AUD_1$, which had a novel visual component but matched the training target's auditory component. Examples of a target, and two choice sets are presented in Figure 1.

Design and Procedure

Child participants were tested in a quiet room within their daycare center. Adult participants were tested in a small laboratory room on campus. Child

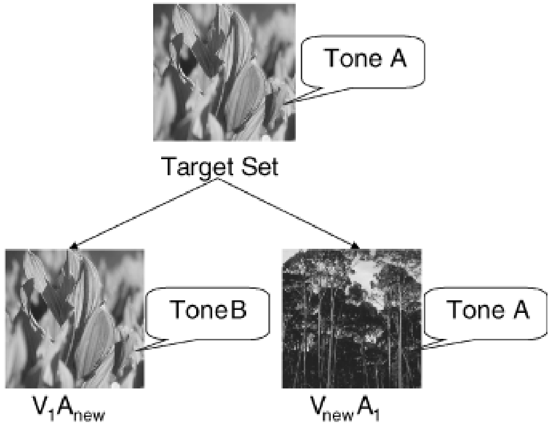

Figure 1. Examples of the stimulus sets used in Experiment 1.

participants were told they would play a game in which they should find the location of a prize and that they would be rewarded at the end of the game with a prize. Game and prize descriptions were omitted for the adult participants.

The overall experiment included six blocks, with each block consisting of eight training trials (a training session) and six test trials (a testing session). Child participants were presented with one to two blocks per day, and the experiment was spread over 2 to 3 weeks. Adult participants completed all six blocks in one sitting. All stimuli were presented on a Dell Inspiron laptop computer, and presentation of stimuli and recording of responses was controlled by a specially developed program written in Visual Basic, which allowed recording of each response and its latency. Adults responded by pressing appropriate buttons on the keyboard, whereas children responded by pointing, and the experimenter pressed appropriate buttons. Because of a high variance in children's latency data, latency data were not analyzed formally. Small toys were given to child participants at the end of each day as reward for participation.

Training session. Stimuli were presented in the following manner. First, either VIS_1 or VIS_2 was presented on one side of the screen, accompanied with either AUD_1 or AUD_2, respectively. Second, the remaining visual stimulus was presented on the other side of the screen, accompanied with the remaining auditory stimulus. The order of appearance and the side of the screen was counterbalanced for the two stimulus sets across training trials, such that each set could appear either first or second and on either the right or left side of the screen. A white circle icon replaced each set at the end of its presentation. The goal of training was to teach the children consistently to select the VIS_1AUD_1 stimulus set; therefore, on each trial children were provided with yes feedback when this stimulus set was chosen and no feedback when the VIS_2AUD_2 stimulus set was chosen. Only participants making correct selections in the final four trials moved into the test session of each eight-trial block. Two of 15 young children did not accomplish a single training session and were eliminated from further analyses.

Test session. During the test session, which immediately followed the training session, participants were presented with two novel stimulus sets. Set VIS_1AUD_{new} matched the training target's visual component but had a novel auditory component, whereas set $VIS_{new}AUD_1$ had a novel visual component but matched the training target's auditory component. Similar to the training session, participants were asked to identify the set where a prize was hidden. Again, the positions of the two stimulus sets were counterbalanced across test trials, and a white circle icon replaced each set at the end of its presentation. When the participant pointed to the selected set, the experimenter pressed the keyboard key corresponding to the selection, without giving feedback to the participant. The overall structure of training and testing trials is presented in Table 1.

Results and Discussion

In the course of training, 13 child participants successfully accomplished 61 of 78 training sessions (78%) and 20 adults successfully accomplished all 120 training sessions. Participants' performance during training trials was not analyzed further, and the analysis focused on the testing phase. Recall that stimulus sets were arranged such that participants could rely either on the visual components of the learned stimulus set (VIS_1) or on the auditory component (AUD_1). Overall means for auditory-based responding were subjected to a one-way ANOVA, with age as a factor, which indicated significant differences across the two age groups (65% in young children vs. 1% in adults), $F(1, 31) = 61.69$, $MSE = 0.07$, $p < .0001$.

Table 1
The Overall Structure of Training and Testing Trials in One Block

Training session ($n = 8$ trials)		Testing session ($n = 6$ trials)	
VIS_1AUD_1 (Trained Set)	VIS_2AUD_2 (Distracter Set)	VIS_1AUD_{new} (Test Set A)	$VIS_{new}AUD_1$ (Test Set B)

Note. VIS = visual stimulus; AUD = auditory stimulus.

To conduct more conservative analyses of participants' performance, we calculated the number of training sessions with above-chance reliance on auditory stimuli, above-chance reliance on visual stimuli, and chance performance. Performance was considered above chance if the same choice was made on five of six trials (binomial test, $p = .09$); otherwise, it was considered at or below chance. In the group of young children the proportion of auditory and visual choices were above chance in 52 of 61 sessions, with the proportion of choices being at or below for all other sessions, whereas in the group of undergraduate students the proportion of choices for all test sessions were above chance. Mean numbers of sessions with above-chance performance by age group and stimulus modality are presented in Figure 2. Percentages of sessions with above-chance auditory choices and above-chance visual choices were subjected to two separate one-way ANOVAs with age as a factor (separate ANOVAs were conducted because auditory and visual choices were nonindependent). There were significant differences across age groups, both Fs(1, 31) > 30, ps < .0001, with young children more likely to rely on auditory stimuli than undergraduate students, and undergraduate students more likely to rely on visual stimuli than young children.

In terms of individual patterns of responses, there emerged three distinct patterns: (a) participants who were above chance in relying on auditory stimuli (auditory responders), (b) participants who were above chance in relying on visual stimuli (visual responders), and (c) participants who were at chance (mixed responders). Above-chance performance was determined by subjecting the total number of auditory and visual choices made by each individual to the binomial test. Percentages of responder types across age groups are presented in Table 2.

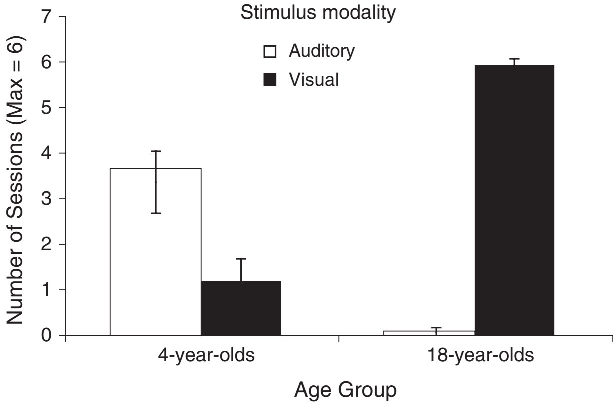

Figure 2. Number of test sessions at above-chance performance by age group and stimulus modality in Experiment 1. Error bars represent standard errors of the mean.

Numbers of auditory, mixed, and visual responders by age group were subjected to a chi-square analysis. The analysis pointed to significant differences among the groups, $\chi^2(2,\ N = 32) = 23.4$, $p < .0001$. The analysis of standardized residuals indicated that auditory responding was the most likely pattern in the group of young children, whereas visual responding was the most likely pattern in the group of undergraduates (all ps < .05).

In short, young children were more likely to rely on the auditory component than on the visual component of the presented stimuli, thus suggesting that the auditory components were more salient for this group than the visual components. This is especially important given that the auditory and the visual components had comparable discriminability (see results of the calibration experiment). Adults, on the other hand, relied almost solely on the visual components.

Experiment 1 leaves several questions unanswered. First, the observed patterns could stem from either the participants' failing to notice changes in either visual or auditory components of the stimuli, or from the participants' deliberate strategy. In particular, if participants were attending to both the auditory and visual stimuli simultaneously, they should notice that they were being forced to choose between the auditory and visual components of the target set. Data provided suggestive evidence that adults exhibited strategic responding, whereas children were more likely to attend to the auditory components than to visual components. In particular, it was observed informally that adults recognized that they were forced to choose between the auditory and visual components, as they were agitated by the choice options, but chose to continue their selections based on the visual components. Children, on the other hand, never seemed to take notice or ask any questions. Because this evidence is suggestive, we deemed it necessary to conduct Experiment 2 examining this issue directly.

Table 2
Percentages of Responder Types by Age Group in Experiment 1

	Responder type		
Age group	Visual	Auditory	Mixed
4-year-old children	15.38	53.83	23.07
Undergraduates	100.00	0.00	0.00

Second, it could be counterargued that the greater salience of auditory stimuli stemmed from young children's inability to meet task demands. In particular, the appearance of the flashing white circle icon could have resulted in masking of the visual stimuli, whereas no such interference would occur for the auditory stimuli. Also, it may be argued that children were unable to process the visual stimuli in such a short time (i.e., 1 s). However, this possibility seems unlikely given that young children had no difficulty processing visual stimuli in the calibration experiment in which the visual stimuli were presented for 1 s as well. To address these concerns, we conducted Experiment 1a, which was similar to Experiment 1 except that auditory components were eliminated from the task. This modification, though reducing the overall task demands (i.e., only visual stimuli were presented in this task), preserved task demands with respect to visual stimuli. If results of Experiment 1 stemmed from the inability to process visual stimuli, young children should perform poorly on Experiment 1a. However, if children could successfully complete the task, their preference for auditory stimuli could not be explained by an inability to process the visual stimuli.

Experiment 1a

Method

Participants

Participants were 10 young children ($M = 3.88$ years, $SD = 1.68$ years; 3 boys and 7 girls) recruited from several daycare centers located in middle class suburbs of Columbus, Ohio. The majority of participants were Caucasian.

Materials and Procedure

Visual stimuli were the same as used in Experiment 1. No auditory stimuli were used in this experiment. The procedure was identical to that of Experiment 1.

Results and Discussion

Participants made correct selections on 83% of all test trials. They successfully completed 18 of 20 test sessions, 72% of which were above chance visually (i.e., the same choice was made on five of six trials; binomial test, $p = .09$). These results indicate that in the absence of auditory components, young children had little difficulty processing visual components. Therefore, it seems that results of Experiment 1 did not stem from task demands, such as masking of the visual components.

Although Experiments 1 and 1a indicate that young children were more likely to rely on visual stimuli than on auditory stimuli, these experiments do not elucidate the processing of auditory and visual stimuli. Recall that patterns of responses observed in Experiment 1 could be indicative of higher salience of either visual or auditory components of the stimuli or they could be indicative of deliberate strategies used by participants. To distinguish between these possibilities we conducted Experiment 2.

Experiment 2

Experiment 2 was designed to examine whether the salience of the nonlinguistic auditory patterns for children and the salience of visual stimuli for adults in Experiment 1 could be explained by strategy factors or by attentional factors. To accomplish this, a same–different immediate recognition task was designed to determine whether both sounds and images were being attended to equally. For example, if the children were capable of attending equally to both sounds and images, the greater weight of sounds found in Experiment 1 could be based on preference, such that they found the tones more interesting or deemed them more important. If children are not capable of equally attending when the specific sounds and images compete (e.g., the child can only attend to the sounds), it may be that when the specific sounds and images are presented simultaneously, the child can only attend to one type of stimuli, and therefore, sounds have greater weight because they truly overshadowed the visual stimuli.

Method

Participants

Participants were 15 young children ($M = 4.19$ years, $SD = 0.41$ years; 6 girls and 9 boys) recruited from several daycare centers located in middle class suburbs of Columbus, Ohio, and 15 college undergraduates ($M = 21.11$years; $SD = 2.98$; 3 women and 12 men) from a large midwestern university participating in the experiment in partial fulfillment of a course requirement. The majority of participants were Caucasian.

Materials

The same auditory and visual stimuli used in Experiment 1 were used in the current experiment.

Recall that children were above chance in attending to and correctly discriminating both the visual and auditory stimuli when either modality was presented alone in the same–different immediate recognition task used in the calibration study in Experiment 1. In Experiment 2, these stimuli were presented in sets composed of a simultaneous presentation of visual and auditory component, such that each image's presentation matched the duration of the accompanying sound. The six VIS$_1$AUD$_1$ sets from Experiment 1 were used as targets. Each target was followed by a test (i.e., recognition) set, and there were four types of test sets: (a) test set that contained the same auditory and same visual components as one of the target sets (i.e., VIS$_1$AUD$_1$), (b) test set that contained an auditory component and a visual component different from one of the target sets (i.e., VIS$_{\text{new}}$AUD$_{\text{new}}$), (c) test set that matched a target set's visual component but had a novel auditory component (i.e., VIS$_1$AUD$_{\text{new}}$), and (d) test set that had a novel visual component but matched a target set's auditory component (i.e., VIS$_{\text{new}}$AUD$_1$). These test sets constituted four within-participants recognition conditions (see Figure 3 for an example for each of the four conditions).

Design and Procedure

Child participants were tested individually in a quiet room within their daycare centers. Adult participants were tested in a small laboratory room on campus. Each participant entered the room and sat in a chair in front of a laptop computer. Participants were told they would play a matching game in which they would be shown one item, followed by a second item, and their goal would be to identify whether the two items were "the same" or "not the same." The experiment included 24 trials. In each trial, a target set was presented and followed by a test set, and participants were then prompted to respond whether the test set was same or different as the target set. Stimuli were displayed in the following manner. The target set, VIS$_1$AUD$_1$, was presented on the center of the screen for 1 s, followed by a blank screen for 1 s. Next, a second stimulus set (i.e., one of the four test items) was presented on the center of the screen followed by a pause until the participant responded. Stimuli were presented and responses were recorded on a Dell Inspiron laptop computer, and presentation of stimuli and recording of responses was controlled by Superlab Pro 2.0 software (Cedrus Corporation, 1999). The presentation order of the six targets and their four conditions was randomized. Children were given small toys at the end of the experiment as rewards for their participation.

Results and Discussion

Recall that the test sets were arranged into one of four conditions: (a) VIS$_1$AUD$_1$, (b) VIS$_{\text{new}}$AUD$_{\text{new}}$, (c) VIS$_1$AUD$_{\text{new}}$, and (d) VIS$_{\text{new}}$AUD$_1$. Data were analyzed to determine whether children were equally capable of identifying differences in both auditory and visual stimuli; therefore, the data from the VIS$_1$AUD$_1$ and VIS$_{\text{new}}$AUD$_{\text{new}}$ conditions served as controls, and the data from the VIS$_1$AUD$_{\text{new}}$ and VIS$_{\text{new}}$AUD$_1$ were of major interest. As shown in Figure 4, children were above chance in correctly identifying both controls (more than 88% correct in VIS$_1$AUD$_1$ and VIS$_{\text{new}}$AUD$_{\text{new}}$ conditions), both one-sample $ts > 5.8$, $ps < .0001$. This high accuracy indicates that children took the task seriously and attended to the trained stimuli. They also exhibited

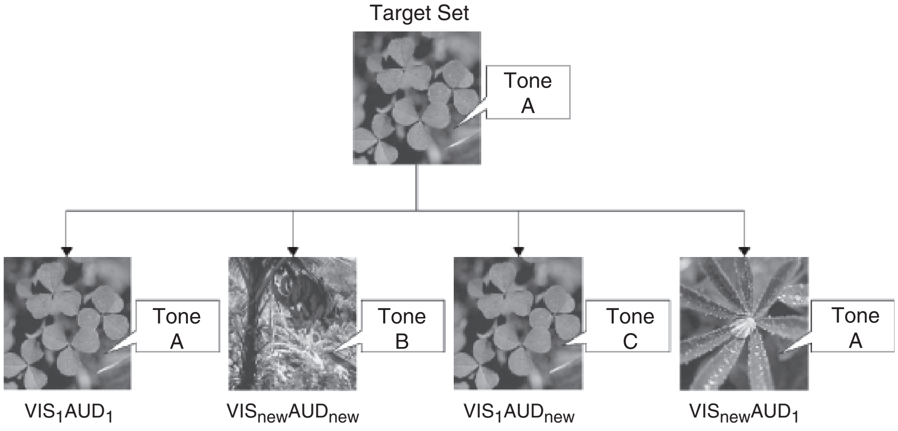

Figure 3. Examples of the stimulus sets used in Experiment 2.

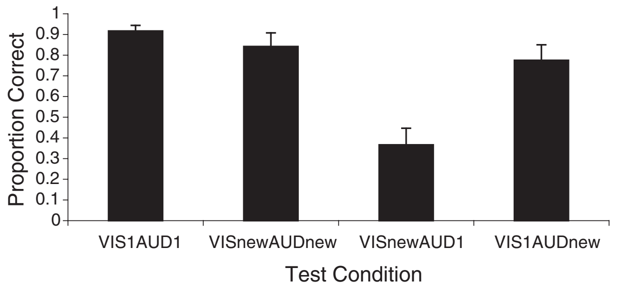

Figure 4. Proportions of correct same/different responses by child participants in Experiment 2. Error bars represent standard errors of the mean.

above-chance accuracy (80% correct) in rejecting sets that had changed auditory components with the same visual components as the targets (VIS_1AUD_{new}), one-sample $t(14) = 4$, $p < .005$, whereas they exhibited below-chance accuracy (30% correct) in rejecting sets that had changed visual components but shared auditory components with the targets ($VIS_{new}AUD_1$), one-sample $t(14) = –2.5$, $p < .05$. At the same time, adults made no errors in all four conditions, thus exhibiting greater accuracy than children.

Overall, results indicate that children are more likely to attend to auditory stimuli than to visual stimuli. Although the results of Experiment 1 and 2 support our predictions that auditory stimuli may have greater attentional weights for young children than visual stimuli, it is possible that these results are stimuli specific. Even though young children can discriminate the visual images when necessary (as observed in the calibration experiment and Experiment 1a), visual stimuli could be more complex and more difficult to process than auditory stimuli. Perhaps if we had used visual stimuli composed of simpler features, the auditory preference would be less robust. To determine whether the results of the reported experiments could be generalized to other types of visual stimuli, we replicated Experiments 1and 2 with stimuli composed of simple geometric objects.

Experiment 3

Method

Participants

Participants were 16 young children (8 boys and 8 girls) who ranged between 4 and 5.08 years in age ($M = 4.41$ years, $SD = 0.3$ years) recruited from several daycare centers located in middle-class suburbs of Columbus, Ohio. The majority of participants were Caucasian.

Materials

Sixteen stimulus sets were created. Each set was composed of a simultaneous auditory and visual component. Each visual component included three green simple geometric shapes (e.g., triangles, diamonds, squares, circles, crosses, and octagons) that were presented in a horizontal line. Six shapes were used, and in each pattern the shapes and their order were randomly selected. Each individual shape was 2.5 cm × 2.5 cm, and the total pattern was enclosed by a rectangle border that was 10 cm × 5 cm. Auditory stimuli were the same as used in Experiment 1. A calibration experiment was performed to ensure that the visual stimuli were sufficiently discriminable (note that it was previously determined that the auditory stimuli were discriminable). This was accomplished using the same–different immediate recognition task identical to the one used for the calibration experiment in Experiment 1. A different sample of 14 children 4 years of age exhibited near ceiling performance, discriminating the new visual stimuli in 88% of trials, which was above chance, $t(12) > 15$, $p < .0001$.

Design and Procedure

The procedure was similar to Experiment 1 except that children were tested with two separate blocks presented in one sitting (as opposed to six blocks presented over two sittings in Experiment 1).

Results and Discussion

Participants made auditory-based choices on 71% of trials, which was above chance, one-sample $t(12) = 2.98$, $p < .05$. Of 32 sessions, participants successfully completed 24 sessions, of which 54% favored the auditory modality, 17% favored the visual modality, and 29% were at chance. A chi-square analysis pointed to a significant difference among response patterns, $\chi^2(2, N = 72) = 7.88$, $p < .05$. The analysis of standardized residuals indicated that auditory responding was the most likely pattern of response, $p < .05$. In short, the results of this experiment replicated the results of Experiment 1, indicating that participants' responses in Experiment 1 were not stimuli specific.

Experiment 4

Experiment 4 was designed to examine whether the salience of the visual patterns used in Experiment 3 was attributable to preferential factors or attentional factors. Recall that in Experiment 2 it was demonstrated that children were only attending to the auditory stimuli. To determine whether overshadowing of visual by auditory was also occurring in Experiment 3, the same–different immediate recognition task used in Experiment 2 was repeated with the stimuli used in Experiment 3.

Method

Participants

Participants were 15 young children ($M = 4.37$ years, $SD = 0.24$ years; 7 girls and 8 boys) recruited from several daycare centers located in middle-class suburbs of Columbus, Ohio. The majority of participants were Caucasian.

Materials

The same auditory and visual stimuli used in Experiment 3 were used in this experiment. Similar to Experiment 2, visual and auditory components were presented simultaneously (i.e., each image's presentation matched the duration of its sound). As in Experiment 2, four types of recognition items were created: (a) items that contained the same auditory and same visual components as one of the target items (i.e., VIS$_1$AUD$_1$), (b) items that contained auditory and visual components different from one of the target items (i.e., VIS$_{\text{new}}$AUD$_{\text{new}}$), (c) items that matched a target set's visual component but had a novel auditory component (i.e., VIS$_1$AUD$_{\text{new}}$), and (d) items that had a novel visual component but matched a target set's auditory component (i.e., VIS$_{\text{new}}$AUD$_1$).

Design and Procedure

Design and procedure were identical to those in Experiment 2.

Results and Discussion

As in Experiment 2, the data from the VIS$_1$AUD$_1$ and VIS$_{\text{new}}$AUD$_{\text{new}}$ conditions served as controls, and the data on VIS$_1$AUD$_{\text{new}}$ and VIS$_{\text{new}}$AUD$_1$ conditions were of major interest. As illustrated in Figure 5, results were similar to Experiment 2. First, children exhibited above-chance accuracy in correctly identifying both controls as old and new, respectively, both one-sample ts(14) > 7.3, ps < .0001. More important, they exhibited above-chance accuracy in rejecting VIS$_1$AUD$_{\text{new}}$—the sets that had just the auditory component changed—one-sample $t(14) = 5.4$, $p < .0001$, and below-chance accuracy in rejecting VIS$_{\text{new}}$AUD$_1$—the sets that had just the visual component changed—$t(14) = -2.14$, $p = .05$. Overall, the results of this experiment further corroborate the results of Experiment 2 indicating that across different stimuli young children are more likely to attend to auditory stimuli than to visual stimuli.

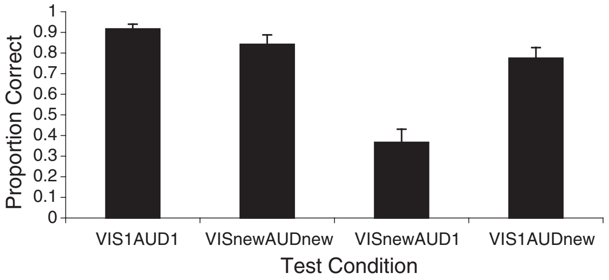

Figure 5. Proportions of correct same/different responses by participants in Experiment 4. Error bars represent standard errors of the mean.

General Discussion

At the most general level, we found that young children made equivalence judgments based on the equivalence of the auditory and not the visual component (Experiments 1 and 3) and that they more readily encoded the auditory components than the visual components (Experiments 2 and 4). These results indicate that across various visual stimuli young children exhibit dominance of the auditory modality over the visual modality. The results of the control and calibration experiments (i.e., Experiment 1a and calibration studies reported in Experiments 1 and 3) demonstrated that in the absence of auditory stimuli, participants had no difficulty processing visual stimuli. Therefore, processing of visual stimuli was not difficult per se, but rather it was mediated by the presence or absence of auditory stimuli. When both visual and auditory stimuli were presented simultaneously, children were more likely to attend to auditory stimuli than to visual stimuli.

Although the dominance of the auditory modality has been well established in 6- to 10-month-old infants (Lewkowicz, 1994), the dominance of auditory modality in young children is a novel finding. It has been generally accepted that the dominance of the auditory modality disappears by late infancy (see Lewkowicz, 1994, for a review), but there has been little research on the auditory dominance in young children. Current research fills this gap indicating that the dominance of the auditory modality found in infants continues well into the preschool years.

Of course, it is well established that humans are flexible attenders, and under different conditions they attend to different properties of stimuli (Jones & Smith, 2002; Jones, Smith, & Landau, 1991; Nosofsky, 1986; Smith, Jones, & Landau, 1996). Therefore, it is possible that under some stimuli or task conditions young children may attend to visual stimuli even when auditory stimuli are presented. However, the reported research has shown that, in general, young children (unlike adults) are strongly biased to pay more attention to the auditory modality.

These results have important implications for our understanding of the prominence of linguistic labels for categorization, induction, and similarity judgment. Two classes of explanations have been proffered: one arguing for language-specific effects (prosodic, semantic, or both) and another arguing for general auditory effects. Present findings indicate that the prominent role of linguistic labels may be explained in part by the general auditory effects. Of course, the support of the general auditory explanation does not rule out the language-specific explanations, and the size of language-specific effects will be examined in our future research.

Why do auditory stimuli dominate visual stimuli for infants and young children, and why does this dominance disappear with age? Although an empirical examination of these reasons is outside the scope of the present research, we consider several possibilities. First, it is known that the auditory system matures earlier than the visual system: The auditory system starts functioning during the last trimester of gestation (Birnholz & Benaceraff, 1983; Gottlieb, 1971; see also Jusczyk, 1998, for a review), whereas the visual system does not start functioning until after birth. As a result, even though the neural bases of visual perception are fully developed at a young age (e.g., Aslin & Smith, 1988), the visual system may still lag behind the auditory system throughout the early years. However, it is not very likely that this maturational asynchrony underlies the auditory dominance in 4-year-olds found in this research.

Second, it is possible that maturational differences between the auditory and visual modalities peaks when the course of language acquisition is at its height. This privileged status of the auditory modality may be functionally important for language acquisition, and its advantage may start decreasing when the child has (in principle) mastered the task of acquiring language. In this case, young children participating in this research may be at the endpoint of this developmental asynchrony.

Third, it is possible that the dominance of the auditory modality stems from different attentional demands for processing visual and auditory stimuli. Typically, a sound disappears after a relatively short duration, whereas a corresponding visual scene may be present for a much longer duration (e.g., animal's call vs. its appearance), and therefore it seems more adaptive to allocate attentional resources to sounds before allocating them to corresponding visual scenes. In this case, the auditory dominance may decrease with an increase of attentional resources.

Finally, it is possible that adults have a visual bias stemming from their knowledge that visually presented entities are likely to be objects, whereas auditorily presented entities are likely to be events. If adults interpreted reference to stimuli presented in Experiment 1 as objects, this interpretation may have biased them toward visual stimuli. At the same time, children may not have this bias. However, the fact that adults were not biased toward visual stimuli in the recognition task in Experiment 2 (i.e., they encoded auditory and visual stimuli equally well) casts doubt on this possibility.

These as well as other potential explanations have different implications for our understanding of modality dominance, and clarification of mechanisms underlying this dominance requires more theoretical as well as empirical work. Although precise mechanisms underlying the reported privileged processing of auditory stimuli remain unclear, this privileged processing appears to be adaptive. Because visual processing is massively parallel and auditory processing is largely serial, in the absence of privileged processing of auditory stimuli, the visual modality would completely dominate the auditory modality, thus making the task of language acquisition very difficult, if not impossible. The privileged processing of auditory stimuli allows young children not to overlook the auditory input. This ability is specifically important during language acquisition when much of critically important information is presented auditorily.

A privileged processing of auditory stimuli may reflect a more general mechanism of dominance of specific sensory systems that has been found across various species. In particular, research on superior colliculus (SC), the brain area known for its role in initiating orienting behaviors, has demonstrated that SC in different species may be dominated by different modality inputs, although specific dominant modalities may differ across species. For example, SC of primates is dominated by inputs from vision and audition, whereas SC of the rodent family is dominated by touch and nociception (Meredith, Clemo, & Dehneri, 2000). In other words, because it would be detrimental to a species' survival to receive all possible sensory information, the nervous systems of different species are specialized to have dominant sensory inputs, and because survival needs differ across species and across points of development, so does the dominant modality. Thus, the dominance of the auditory modality in the early years may reflect this adaptive mechanism, and examination of the modality dominance in young and mature nonhuman primates may shed light on this mechanism.

We also deem it necessary to consider several alternative explanations of the reported findings. First, it could be argued that the reported effects stem from task demands. Visual stimuli, including the geometric figures used in Experiments 3 and 4, might be too complex and abstract for young children. However, results of Experiment 1a demonstrated that young children had no difficulty performing the task in the absence of auditory components. Furthermore, calibration studies reported in Experiments 1 and 3 indicated that, when auditory stimuli were absent, young children had no difficulty encoding visual stimuli. Therefore, in the absence of auditory stimuli, visual stimuli were not difficult for young children, and thus it is unlikely that the reported dominance of auditory stimuli can be accounted by the difficulty of processing of visual stimuli.

Another possible alternative explanation is that reported results stem from familiarity effects. It is possible that for adults visual stimuli are familiar and auditory stimuli are novel, whereas for young children both visual and auditory stimuli are novel. However, the possibility of differential novelty of stimuli for adults is unlikely. Adults encoded equally well both visual and auditory components. In addition, in both calibration experiments (reported in Experiments 1 and 3), adults' correct responses to auditory stimuli were slightly but reliably faster than their correct responses to visual stimuli.

Another variant of the familiarity explanation could argue that although both visual and auditory stimuli were equally familiar for adults, for young children, auditory stimuli were more familiar than visual stimuli. However, this alternative is inconsistent with both calibration experiments. When visual and auditory stimuli were presented separately, young children encoded both sets of stimuli equally well.

Finally, it could be argued that the reported effects are limited to a set of simple and meaningless sounds. In particular, if presented with more interesting and meaningful sounds (e.g., animal calls), even adults may rely on sounds. Although this possibility should be addressed in future research, it does not undermine the current findings that young children are strongly biased to attend to auditorily presented stimuli.

The reported effects were robust and reliable; however, there are several issues potentially limiting the generalizability of present findings. First, although most research demonstrating facilitative effects of linguistic labels focused on natural kinds (e.g., animals or flowers) or human-made artifacts, the present research focused on more complex and less individualizable entities (i.e., landscapes) or on more simple entities (i.e., arrangements of geometric shapes). Therefore, it could be argued that these findings could not be generalized to natural kinds or artifacts. Second, it could be argued that although a short (1 s long) presentation is typical for auditory stimuli, such short presentation may not be typical for presentation of visual stimuli. Third, it could be argued that, because the auditory stimuli were limited to simple three-sound sequences, the reported dominance of the auditory modality in young children may be specific to these simple meaningless sounds.

Although the first issue was not addressed in the present research, it seems that the auditory dominance is a basic mechanism that should not be sensitive to high-level semantic aspects of visual stimuli. On the other hand, it is possible that those entities that belong to easily namable familiar kinds (e.g., animal or flower) would be processed faster and encoded more ably. Therefore, additional research is needed to examine auditory dominance with more specific, familiar, and easily namable stimuli. In this case, the dominance of the auditory modality would be limited to relatively novel visual stimuli.

The second issue was addressed in this research. First, as demonstrated in the calibration experiments, the 1-s presentation was sufficient to accurately discriminate between visual stimuli. Second, when auditory stimuli were eliminated and participants were presented only with visual stimuli, they had no difficulty processing those stimuli (Experiment 1a). Therefore, the reported results stem not from an inability to process visual stimuli under these conditions but from the dominance of the auditory input.

Finally, the third issue requires additional research with a wide variety of auditory stimuli. However, although it is easy to envision auditory stimuli that would lead to even larger auditory effects (e.g., animal calls, melodies, or sounds of mechanisms), it is difficult to envision auditory stimuli that would lead to smaller auditory effects given the visual stimuli used in this research. Thus, unless the sound level drops significantly below the range of the sound level of human voice in a regular conversation, different auditory stimuli are likely to lead to larger rather than smaller effects of auditory stimuli.

Although several issues require further research, the reported studies point to the dominance of the auditory modality over the visual modality in young children. This is a novel finding (earlier findings of the auditory dominance have been limited to infants) that may elucidate the development of processing as well as shed light on some of the facilitative effects of linguistic labels on similarity judgment, induction, and categorization.

References

Aslin, R., & Smith, L. (1988). Perceptual development. *Annual Review of Psychology, 39,* 435–474.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology, 64,* 3–26.

Birnholz, J. C., & Benaceraff, B. B. (1983). The development of human fetal hearing. *Science, 222,* 516–518.

Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (146–196). New York: Cambridge University Press.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition, 23,* 183–209.

Gottlieb, G. (1971). *Development of species identification in birds: An inquiry into the prenatal determinants of perception.* Chicago: University of Chicago Press.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science, 5,* 219–232.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development, 62,* 499–516.

Jusczyk, P. W. (1998). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Lewkowicz, D. J. (1988a). Sensory dominance in infants: 1. Six-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24,* 155–171.

Lewkowicz, D. J. (1988b). Sensory dominance in infants: 2. Ten-month-old infants' response to auditory-visual compounds. *Developmental Psychology, 24,* 172–182.

Lewkowicz, D. J. (1994). The development of intersensory perception in human infants. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of perception: Comparative perspectives* (165–203). Hillsdale, NJ: Erlbaum.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Meredith, M. A., Clemo, H. R. H., & Dehneri, L. R. (2002). Responses to innocuous, but not noxious, somatosensory stimulation in the ferret superior colliculus. *Somatosensory and Motor Research, 17,* 297–308.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Roberts, K., & Jacob, M. (1991). Linguistic versus attentional influences on nonlinguistic categorization in 15-month-old infants. *Cognitive Development, 6,* 355–375.

Sloutsky, V. M., & Fisher, A. V. (2001). Effects of linguistic and perceptual information on categorization in young children. In J. Moore & K. Stenning (Eds.), *Proceedings of the XXIII Annual Conference of the Cognitive Science Society* (pp. 946–951). Mahwah, NJ: Erlbaum.

Sloutsky, V. M., & Lo, Y. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology, 6,* 1478–1492.

Sloutsky, V. M., Lo, Y., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72,* 1695–1709.

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

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29,* 557–302.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology, 34,* 1289–1309.

Woodward, A. L., & Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development, 70,* 65–77.