CHILD DEVELOPMENT

Child Development, March/April 2010, Volume 81, Number 2, Pages 669–679

Children Monitor Individuals' Expertise for Word Learning

David M. Sobel
Brown University

Kathleen H. Corriveau
Harvard University

Two experiments examined preschoolers' ability to learn novel words using others' expertise about objects' nonobvious properties. In Experiment 1, 4-year-olds (n = 24) endorsed individuals' labels for objects based on their differing causal knowledge about those objects. Experiment 2 examined the robustness of this inference and its development. Four-year-olds (n = 40) endorsed labels from confederates who accurately predicted objects' nonobvious internal properties but not nonobvious external properties. Three-year-olds (n = 40) performed at chance levels in both cases and were less likely to recognize the informants' expertise, suggesting that they might be unable to monitor individuals' expertise. These data suggest that children's ability to learn from testimony is necessary for their understanding of the relevance of an individual's expertise.

As adults, we rely on others' knowledge to learn and to make inferences about the world. We count on mechanics to diagnose and repair our cars because they know more about the functions of fuel-injection systems and carburetors. We allow doctors to advise us on medical decisions because they know more about how diseases cause symptoms and are affected by treatments. We hope our political leaders can effect change in social issues, the economy, or foreign affairs because of their greater experience or deeper knowledge of the issues involved. Whether trust is warranted in a mechanic, doctor, or politician depends on our belief that they understand the causal structures critical to a specific knowledge base. This article concerns whether and at what age children can use an individual's causal expertise to make novel inferences. Specifically, we focus on the effect observing another's causal knowledge has on children's subsequent willingness to rely on that person when learning new words.

The importance of ''trust'' in word learning has been studied extensively because there is a conventional, but arbitrary, assignment of lexical items to concepts and objects (e.g., Bloom, 2000; de Saussure, 1966; see Koenig & Harris, 2008, for a review). As such, children must rely on data they observe to learn the meaning of words. This involves recognizing that others are reliable sources of information. Thus, children should use data gained from a reliable source while discounting information that is unreliable. A growing body of research indicates that by the age of 4 (and possibly younger), preschoolers are able to use a speaker's past history of reliability to make inferences about the extension of novel labels to novel objects (e.g., Clément, Koenig, & Harris, 2004; Corriveau & Harris, 2009; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007). Four-year-olds also assume that adults tend to be reliable by default (e.g., Corriveau, Meints, & Harris, 2009) as well as register the role of uncertainty in word learning (e.g., Sabbagh & Baldwin, 2001). These data all suggest that children trust adults to provide them with accurate lexical information and discount that information when the adult is a less reliable source (but see Nurmsoo & Robinson, 2009, for a different opinion).

Such inferential ability potentially allows children to develop concepts of expertise. Children develop ontological commitments about other kinds of events they cannot directly observe based on the ''patterns of testimony'' that they receive (Harris & Koenig, 2006; Harris, Pasquini, Duke, Asscher, & Pons, 2006). More generally, Keil and colleagues have suggested that young children learn that individuals' knowledge forms a ''division

This research was supported by NSF (DLS-0518161 to DMS). We would like to thank Sheridan Brett, David Buchanan, Silvia Chen, Jon Cohen, Angie Kim, Karina Ikesoe, Emily Myers, Rachel Shelley-Abrahamson, Cesalie Stepney, Elena Tenenbaum, and Lea Ventura for help with data collection and analysis.

Correspondence concerning this article should be addressed to David M. Sobel, Department of Cognitive and Linguistic Sciences, Box 1978, Brown University, Providence, RI 02912. Electronic mail may be sent to dave_sobel@brown.edu.

of cognitive labor'' (Danovitch & Keil, 2004; Keil, Stein, Webb, Billings, & Rozenblit, 2008; Lutz & Keil, 2002). At relatively young ages, children understand that individuals possess different pieces of knowledge, which potentially cohere around causal principles. Keil et al. (2008), in particular, demonstrated that 5-year-olds recognize that others' knowledge is organized around disciplines that reflect the structure of modern universities. Children understood that someone who knows about a physical property is more likely to know about a chemical one (grouping two natural sciences) than a principle of cognitive psychology (grouping a natural science with a social science).

Understanding this ''division of cognitive labor'' requires that children have access to and can make inferences about their own knowledge. Our goal here is not to investigate the emergence of this metacognitive understanding but rather to consider whether children possess the cognitive ability to monitor individuals as reliable sources of knowledge within a domain. This might be the basis on which the metacognitive division of cognitive labor is built. In particular, we are interested in the consistency between individuals' expertise and the inference children are asked to make based on that expertise. For example, several researchers have found that preschoolers rely on speakers' history of accurately labeling familiar objects when those speakers subsequently provided conflicting information about a novel object's function (Birch, Vauthier, & Bloom, 2008; Koenig & Harris, 2005). Using the more reliable speaker in this circumstance seems logical given the coherence between artifacts' category membership and their functions. Preschoolers often relate these two pieces of information, making inferences about one based on the other (e.g., Bloom, 1996; Gelman & Bloom, 2000; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson et al., 1995). The reliability with which an individual categorizes objects should inform children that the individual possesses categorical knowledge about those objects, such as their function. But an individual's history of reliabilty should only be applicable if that reliability is relevant to the inference at hand—as reliable as a plumber might be at fixing a toilet, we might choose to discount his political savvy.

These insights motivated this study. In Experiment 1, we examined whether 4-year-olds could recognize that two individuals were experts in different pieces of causal knowledge. We asked whether 4-year-olds could subsequently use the individuals' expertise when learning the label of a novel object that exhibited a causal property related to only one of the individual's expertise. Several investigations have shown that young children relate objects' causal properties with their category membership, particularly when assigning labels to those objects (Booth & Waxman, 2002; Gopnik & Sobel, 2000; Graham, Kilbreath, & Welder, 2004; Nazzi & Gopnik, 2000). Although some have argued that object labels do not necessarily relate to conceptually deep properties of those objects, like their kind or function (e.g., Smith, Jones, & Landau, 1996), we suggest that children recognize that object labels relate to the conceptual structure indicated by those labels. Thus, we hypothesized that young children would rely on individuals' different causal knowledge judiciously to make inferences about novel objects' labels.

Experiment 2 had three related goals. The first was to examine whether preschoolers used informants' reliability about a different property of objects—its insides or material kind—to make inferences about novel object labels. Several investigations have demonstrated that 4-year-olds, but not 3-year-olds, understand that objects' causal properties or category membership are related to those objects' insides (Gelman & Gottfried, 1996; Gelman & Wellman, 1991; Gottfried & Gelman, 2005; Sobel, Yoachim, Gopnik, Meltzoff, & Blumenthal, 2007). Given this developmental difference, we hypothesized that 4-year-olds, but not 3-year-olds, would relate individuals' accurate histories of describing nonobvious internal properties of objects with their category membership.

But, an individual's reliabilty should only be applicable if that reliability is relevant to the inference at hand. The second goal of Experiment 2 was to consider whether children treated any set of reliable actions as relevant to word learning or whether children distinguished between relevant and irrelevant training. To investigate this possibility, we presented another group of 3- and 4-year-olds with informants who possessed different knowledge about an external property of a set of objects. Half of the objects had green stickers on them and half had red stickers. One informant was accurate in identifying which objects had red stickers on them but was ignorant about the other objects; the other informant was accurate in identifying the objects with green stickers on them but was ignorant about the objects with red stickers. Our hypothesis was that this expertise would be irrelevant to the objects' category membership: Neither age group should use the speakers' reliability about these external parts to make inferences about category membership. Furthermore, the 4-year-olds in particular should be more likely to use the speakers' reliability in the insides condition than in the outsides condition.

Experiment 1

Four-year-olds were introduced to two individuals who demonstrated different pieces of causal knowledge about the way in which objects activated a machine. We then showed children a new object whose efficacy on the machine was consistent with the predictions of one of the two individuals. We examined whether children would use the individuals' different knowledge to decide which informant to turn to when making an inference about the label of the new object.

Method

Participants

Twenty-four 4-year-olds (16 boys and 8 girls, *M* = 53.71 months, *SD* = 2.90) were recruited from a preschool classroom, fliers posted at local preschools, and a list of hospital births. Four additional children were tested but not included in the final analysis: 2 because of experimental error and 2 failed to interact with the experimenter. Most children were Caucasian and from middle- to upper-middle-class families; however, no formal measure of socioeconomic status (SES) was taken. The participants did represent the diversity of the population from which they were sampled.

Materials

The machine was an 8 × 6 × 3 in. box made of black plastic with a white top (see Figure 1). If an object was placed on it, the machine would activate, both by illuminating either red or green light-emitting diodes (LEDs), which were wired into the top of the device, and by playing one of two different musical arrangements. The machine worked via a remote control (hidden from the child), which the experimenter used to control whether the machine activated red or green. All of the objects used activated the machine, but the remote could also ''enable'' or ''disable'' the machine, such that other objects that a child might inadvertently put on the machine (e.g., fingers) would not activate it. Eight wooden blocks,

Figure 1. The machine used in the Experiment 1.
Note. The machine would activate by illuminating red or green light-emitting diodes (LEDs) wired into the white top. When the machine activated, it also played one of two songs, dependent on whether it was set to light up red or green. The objects used in Experiment 1 were all wooden blocks without any distinguishable features. The object shown in Figure 1 is an object used in Experiment 2, with a dowel over a cavity, revealing its ''inside'' (in the insides condition) and with a sticker on the back (not shown), both of which were either red or green.

all of different sizes, colors, and shapes were used. No block was red or green.

Procedure

Children were tested in a quiet room at their school or in the lab. The child and experimenter were seated across from each other at a table. The experimenter told the child that they would play a game with two of the experimenter's friends. The two confederates then entered the room and sat on either side of the experimenter. The confederates were always the same gender and never wore red or green clothing. The child was then told that they would all play a game together with the experimenter's machine, which was brought out onto the table. The experimenter said, ''This is a special machine. I have some toys. Some toys make the machine go red and some toys make the machine go green, and we're going to try and figure it out. Let's take a look.''

Training trials. Children observed six training trials. In each, the experimenter brought out a block (randomly chosen from the set of eight) from under the table. On the first training trial, the experimenter told the child, ''Take a look at this toy. I don't know what's going to happen when we put this on the machine, but my friends are pretty smart, so let's ask them.'' (On subsequent trials, the experimenter just showed the child the new object and said, ''Let's ask my friends what this one does.'')

The experimenter turned to one of the confederates (counterbalanced across children) and asked ''What do you think will happen when we put this on the machine?'' Each confederate was given a chance to respond. Prior to the start of the procedure (and unbeknownst to the child), one confederate was designated the *red expert* whereas the other confederate was the *green expert*; we randomly determined who the red expert was for each participant. The red expert always predicted when an object would activate the machine red but was ignorant of objects that made the machine go green (and responded ''I don't know'' for those objects). Similarly, the green expert always responded correctly for objects that made the machine go green and was ignorant for objects that made the machine go red.

After each confederate stated her belief about what the object would do, the experimenter repeated their claims and then placed the object on the machine, which activated consistent with the expert's claim. Children were then asked to point to the person who was right (i.e., who correctly predicted the event). Corrective feedback was given if necessary. These questions were included as a catch trial—our intention was to replace the data from any child who answered more than one of these questions incorrectly from the final analysis. However, all children answered this question correctly on at least five of the six trials. On three of the six training trials, the object activated the machine red and, on the other three, the object activated the machine green (order was randomly determined). Thus, children saw three trials in which the red expert was accurate and the green expert was ignorant (all about red activation events), and three trials in which the green expert was accurate and the red expert was ignorant (all about green activation events). Children never observed either confederate respond incorrectly.

Test trials. Immediately after the training trials, children then observed two test trials. In each, the child was shown a new object and was told ''Let's see what this one does on the machine.'' The object was placed on the machine without asking either confederate her belief about its efficacy. On one trial, the object made the machine go red; on the other, it made the machine go green. The experimenter narrated the machine's activation for the child (e.g., ''Look, this one made the machine go green''). The experimenter then said to the child, ''Do you see this toy. I'll tell you a secret. This toy has a special name. I don't know the special name of this toy, but we can ask my friends.'' Children were then asked the *ask* question, ''Who do you want to ask for the special name of the toy, [Confederate 1] or [Confederate 2]?''

After the child responded, the experimenter asked each confederate for the special name of the toy. Each confederate generated a novel label (*dax* vs. *wug* on one trial, *modi* vs. *toma* on the other). Each label was repeated for the child, and then the child was asked the *endorse* question—for example, ''What do you think the special name of this toy is called, a *modi* or a *toma*?''

After these two trials, children were asked two *knowledge* questions, ''Who knows about things that make the machine go red?'' and ''Who knows about things that make the machine go green?'' These questions were asked to ensure that children remembered the expertise of the two confederates. The location of the red and green experts (with respect to the experimenter) and the particular word labels the experts generated were both counterbalanced.

Results

Overall, children required feedback on ~4% of the training trials, and 18 of the 24 children required no feedback at all. This suggested that children understood the training. Children received a score of 1 on each question if they responded consistently with the confederate with the relevant causal expertise (e.g., if the test object activated the machine green, children were given a score of 1 if they asked the green expert for the special name of the toy on the ask question and endorsed that confederate's label on the endorse question). Preliminary analyses suggest that these scores did not differ between the two ask questions, the two endorse questions, or the two knowledge questions, all McNemar $\chi^2(1, n = 24)$ tests = *ns*, so we combined them. These data are shown in Table 1.

We first analyzed responses on the ask and endorse questions. Only responses on the endorse questions showed a pattern significantly different from chance levels, $\chi^2(2, n = 24) = 6.92$, $p = .031$. The distribution of responses to the ask question was not different from chance, $\chi^2(2, n = 24) = 0.92$, *ns*. Responses to neither of these questions significantly correlated with age.

We next examined responses to the knowledge question. The distribution of responses on this question was significantly different from chance,

Table 1
Distribution of Responses to the Ask, Endorse, and Knowledge Questions in Experiment 1

	Number of correct responses				
	0	1	2	*M*	*SD*
Question type					
Ask	5	11	8	1.12	0.74
Endorse	2	11	11	1.38	0.65
Knowledge	1	2	21	1.83	0.48

with 21 out of the 24 children responding correctly on both trials, $\chi^2(2, n = 24) = 50.00, p < .001$. As with the ask and endorse questions, responses did not correlate with the child's age.

Finally, we wanted to be certain that responses to the ask and endorse questions were representative of children who understood the expertise of the informants. We reanalyzed these responses looking only at the children who answered both knowledge questions correctly. These analyses paralleled what was presented earlier: On the endorse question, children used the informants' expertise on 72% of the trials, with a distribution significantly different from chance, $\chi^2(2, n = 21) = 7.72, p = .021$. On the ask question, children only used the informants' expertise on 65% of the trials, with a distribution not significantly different from chance.

Discussion

These data suggest that children understood the expertise of the confederates and that 4-year-olds recognize that individuals have access to different pieces of causal knowledge. These data also suggest that 4-year-olds have some understanding that another's knowledge can be used to make novel inferences. In particular, children can use individuals' reliability about objects' causal properties to make inferences about those objects' labels.

Although the pattern of responses on the ask and endorse questions was not significantly different from each other, the distribution of responses to the ask question was at chance levels, whereas responses to the endorse question were significantly different from chance. We suggest two possible explanations for this difference. First, previous investigations (e.g., Birch et al., 2008; Jaswal & Neely, 2006; Koenig & Harris, 2005) presented children with a consistently accurate informant and a consistently inaccurate informant. Here, children were asked to monitor the accuracy of different kinds of knowledge. It is possible that requiring children to monitor both informants' accuracy taxed their metacognitive abilities, making the ask question slightly more difficult than the endorse question. Consistent with this idea, Pasquini et al. (2007) asked children to monitor the relative accuracy between the two informants. When one informant was always correct or always incorrect, 4-year-olds' responses to the ask and endorse questions were both above chance. But when both informants responded with different degrees of inaccuracy (one was accurate on 75% of the training trials, the other on 25%), 4-year-olds showed a similar pattern of performance on ask and endorse questions to the results presented here.

Second, unlike much previous research on children's understanding of others' reliability, we showed children live confederates instead of asking them to make inferences about people on videos, characters, or puppets (Corriveau et al., 2009, used a similar procedure). Regardless of the child's choice on the ask question, the informants generated their labels in a fixed order. Thus, children sometimes chose to ask one confederate, but the other confederate offered her label first. It is possible that children thought the ask question referred to whom the experimenter would ask first for a label. We changed this aspect of the procedure in Experiment 2.

In summary, these data suggest that 4-year-olds can both extract and use differential expertise about the causal properties of objects. Children can monitor the quality of information that they hear from the confederates in order to predict which of the two has the relevant information about that object's category membership. At issue is how children are monitoring this information. Our hypothesis is that children are not only monitoring the informants' accuracy but are also integrating their representation of those individuals' reliability with their existing knowledge of the relevance of that information. As children develop knowledge about what information is important for category membership, they should use that knowledge to interpret the reliability information they observe. In order to understand that a plumber's ability to label pipe is independent of that person's ability to craft domestic policy, one must recognize that the causal relations involved in unclogging a drain are different from the causal relations involved in balancing the federal budget.

However, an alternative possibility is that in Experiment 1, children might have just associated the two causal properties with the two individual informants. On this view, children endorse the red expert's label for the object that activated the machine red because they have learned to associate that individual with objects that have particular efficacy. This suggests that 4-year-olds might treat reliable answers to any kind of question about the objects as relevant to the inferences they are asked to make. Experiment 2 examines whether 4-year-olds use any kind of expertise as relevant to making inferences about objects' labels or if they limit their use of the informants' expertise to information related to category membership.

Also, what about younger children? Experiment 1 showed that 4-year-olds are capable of monitoring informants' causal knowledge to make inferences about category membership. Experiment 2 examined whether younger children have similar monitoring abilities. Given that Pasquini et al. (2007) found that 3-year-olds tend to be unable to track individuals' reliability when both are stochastic, 3-year-olds might struggle tracking informants' knowledge.

Experiment 2

Three- and 4-year-olds were introduced to confederates who possessed different pieces of knowledge about unfamiliar objects. This knowledge was either relevant or irrelevant to inferences about those objects' category membership. Instead of presenting children with objects that had different causal properties, we relied on another property of objects that has been found related to their category membership: insides. In Experiment 2, half of the children encountered two informants who were experts about the internal properties of unfamiliar objects. In particular, one expert always knew whether an object had red insides but was ignorant about objects with green insides; the second expert knew when objects had green insides but was ignorant about the objects with red insides. The other half of children learned that the informants were knowledgeable about whether those objects had a particular colored sticker on the back (an analogous nonobvious external property). Several investigations have suggested that objects' insides are related to category membership whereas this kind of external property is not (Gelman & Wellman, 1991; Gottfried & Gelman, 2005; Sobel et al., 2007). We examined whether children would only use expertise related to category membership (i.e., insides) to make inferences about whom to ask for information about object labels and whose labels they should endorse.

Previous investigations have suggested that children develop an understanding of the relation between insides and category membership between the ages of 3 and 4 (e.g., Gottfried & Gelman, 2005; Sobel et al., 2007). Other investigations (e.g., Pasquini et al., 2007) found that 3-year-olds struggled to recognize the relative accuracy of two informants, suggesting that they might also struggle at accurately keeping track of the informants' expertise regardless of the relevance of that information. Given these findings, we predicted an interaction between age and expertise, particularly for responses on the endorse questions: Four-year-olds should relate the confederates' knowledge to the objects' labels only when shown that the confederates have expertise about the internal properties of the objects, but not the external properties of the objects. Three-year-olds should not use that knowledge in either condition.

Method

Participants

Forty 3-year-olds (19 girls, M = 42.95 months, SD = 3.46, range = 36–48 months) and forty 4-year-olds (19 girls, M = 55.35 months, SD = 3.40, range = 49–60 months) were recruited from a set of preschools and a list of hospital births. Most children were Caucasian and from middle- to upper-middle-class families; however, no formal measure of SES was taken. The participants did represent the diversity of the population from which they were sampled.

Materials

Six wooden blocks, each different in shape and color, were used. Each block had a 1/2 in. (diameter) hole drilled into it, covered by a wooden dowel. Behind this dowel was either a red or green plastic pin (three of each color), which gave the appearance of an internal part (labeled ''stuff'' in the procedure). Each block also had a 3/4 in. sticker on the back colored red or green; the color of the sticker matched the color of the internal part. None of the blocks was red or green. Children either saw the internal part or the sticker, and they did not know about the other property of the block.

Procedure

As in Experiment 1, the experimenter introduced the child to two confederates. Children were told they would play a game with a set of toys. In the *insides* condition, children were told that some toys were made out of red stuff and some toys were made out of green stuff. In the *outsides* condition, children were told that some toys had red stickers on the back and other toys had green stickers on the back.

Training trials. The structure of the training trials was similar to Experiment 1. Children were told that the ''friends'' knew about the toys, so each was asked what they thought about each toy (i.e., what kind of stuff it was made of, or what color its sticker was) one at a time. The objects always had their door in place and were held and brought out from under the table such that neither the child nor the confederates could see the objects' insides or the stickers. One confederate was always correct about an object containing red stuff or having a red sticker on the back and said ''I don't know'' otherwise. The other confederate was always accurate about an object containing green stuff or having a green sticker on the back and said ''I don't know'' otherwise. Thus, on each trial, children were told what each confederate thought, then were shown either the internal or external property of the object, and then were asked, ''Who was right?'' Corrective feedback was given if children answered incorrectly. There were four training trials, two in which the red confederate was correct and two in which the green confederate was correct (order was randomly determined). This denotes a difference between this experiment and Experiment 1. Because we intended to work with 3-year-olds in this experiment, we were concerned with the length of the procedure, and thus gave children fewer training trials.

Test trials. Immediately after the fourth training trial, two additional blocks were introduced one at a time. Dependent on condition, the child was shown each block's internal or external property. One block had a red inside or outside; the other had a green inside or outside (order counterbalanced). As in Experiment 1, children were told that the objects had special names. On each trial, children were asked the same *ask* question (i.e., ''Who do you want to ask for the special name of this toy?''). After the child responded, both confederates were asked what they thought the special name of the toy was, and both generated novel words (e.g., *dax* vs. *wug*). Children were asked a similar *endorse* question (i.e., ''What do you think the special name of this toy is: a *dax* or a *wug*?''). Finally, after both trials, children were asked two *knowledge* questions—who knew about the toys made up of red or green stuff, or with red or green stickers? As in Experiment 1, the person who served as the red or green expert was determined randomly, and the side of the experimenter on which the red expert sat and the specific words the experts used to label the objects were counterbalanced across children.

Results

Children required feedback on ~4% of the training trials, and 66 of the 80 children required no feedback at all, suggesting that they understood the nature of the training. Responses were scored in the same manner as Experiment 1. Children's scores did not differ between the two ask, the two endorse, or the two knowledge questions (all McNemar χ^2 tests = *ns*). These data were combined and are shown in Table 2.

We first examined responses to the ask questions by analyzing the scores on this question via a 2 (age group) × 2 (condition: insides vs. outsides) analysis of variance (ANOVA). No significant effect of condition was found, but this analysis did find a significant effect of age group with 4-year-olds responding more accurately in general than 3-year-olds, $F(1, 76) = 7.81$, $p = .007$, partial $\eta^2 = .09$. Nonparametric analysis confirmed this result: Four-year-olds were more likely to ask the informant

Table 2
Distribution of Responses to the Ask, Endorse, and Knowledge Questions by Age and Condition (Insides, Outsides) in Experiment 2

	Number of correct responses				
	0	1	2	*M*	*SD*
Insides condition					
3-year-olds					
Ask	6	12	2	0.80	0.62
Endorse	8	9	3	0.75	0.72
Knowledge	6	3	11	1.25	0.91
4-year-olds					
Ask	3	6	11	1.40	0.75
Endorse	2	8	10	1.40	0.68
Knowledge	3	1	16	1.65	0.75
Outsides condition					
3-year-olds					
Ask	3	13	4	1.05	0.61
Endorse	4	13	3	0.95	0.61
Knowledge	9	1	10	1.05	1.00
4-year-olds					
Ask	3	8	9	1.30	0.73
Endorse	6	9	5	0.95	0.76
Knowledge	6	1	13	1.35	0.93

with the appropriate expertise across the conditions, Mann–Whitney $U = 532.00$, $z = -2.81$, $p = .005$, $r = -.31$.

A similar ANOVA performed on responses to the endorse questions revealed a slightly different pattern. Again, no significant effect of condition was found, but a main effect of age group was present with 4-year-olds in general responding more accurately than 3-year-olds, $F(1, 76) = 4.41$, $p = .039$, partial $\eta^2 = .06$. In addition, a significant interaction between age group and condition was found, $F(1, 76) = 4.41$, $p = .039$, partial $\eta^2 = .06$.

Further investigation revealed that 4-year-olds were significantly more accurate on responses to the endorse question than 3-year-olds in the insides condition, $t(38) = 2.94$, $p = .003$ (one-tailed), Cohen's $d = 0.95$, but showed no significant difference in responses to these questions in the outsides condition, $t(38) = 0.00$, *ns*. Four-year-olds were also more accurate in the insides than the outsides condition, $t(38) = 1.97$, $p = .028$ (one-tailed), Cohen's $d = 0.64$, whereas 3-year-olds' responses were not significantly different between these conditions, $t(38) = 0.35$, *ns*. Nonparametric analyses confirmed these results: Four-year-olds were more likely to use the informants' expertise than 3-year-olds in the insides condition, Mann–Whitney $U = 107.00$, $z = 2.69$, $p = .005$ (one-tailed), $r = .43$, but not the outsides condition, Mann–Whitney $U = 199.00$, $z = 0.03$, *ns*. Similarly, 4-year-olds were more accurate in the insides condition than the outsides condition, Mann–Whitney $U = 135.00$, $z = 1.89$, $p = .040$ (one-tailed), $r = .30$, whereas 3-year-olds showed no difference in responses between these conditions, Mann–Whitney $U = 166.00$, $z = 1.02$, *ns*.

We then examined each group's distribution of responses compared with chance. Only 4-year-olds in the insides condition generated distributions of responses different from chance, $\chi^2(2, n = 20) = 9.60$, $p = .008$, for the ask questions and $\chi^2(2, n = 20) = 7.20$, $p = .027$, for the endorse questions, all other $\chi^2(2, n = 20)$ values < 4.41, *ns*.

Four-year-olds were more likely to respond accurately to the knowledge question than 3-year-olds. However, this difference was only marginally significant, $t(78) = 1.74$, $p = .086$, Cohen's $d = 0.39$. This suggests that 4-year-olds might have been better able to monitor the two informants' knowledge. To ensure that the older children's above-chance performance in the insides condition was not just due to this group recognizing who had causal expertise, we explored each group's distribution of responses to the knowledge question. Both age groups responded differently from chance across both conditions on these questions, all $\chi^2(2, n = 20)$ values > 12.29, all p-values $< .01$. Inspection of Table 2 shows that whereas most 4-year-olds answered these questions accurately, 3-year-olds responses were more bimodally distributed. Thus, we were concerned that the younger children did not understand the expertise information available to them in the training.

As a result, we reanalyzed responses from only the children who responded correctly on the two knowledge questions. This analysis was similar to the analyses presented earlier: Four-year-olds in the insides condition showed a pattern of response significantly different from chance on the endorse questions, $\chi^2(2, n = 16) = 6.38$, $p = .041$. On the ask questions, these children's responses were only different from chance at a marginally significant level, $\chi^2(2, n = 16) = 5.50$, $p = .064$.

Discussion

Experiment 2 found that 4-year-olds could monitor informants' expertise about both internal and external properties of objects, but they used this information judiciously. They only asked for and endorsed the labels generated by the insides experts. Three-year-olds had slightly more difficulty monitoring informants' expertise. Nevertheless, even the 3-year-olds who did so successfully did not make the same inferences as the older children, particularly in the insides condition.

Why did 4-year-olds use the insides experts as sources of knowledge for the objects' labels, but not the outsides experts? In Experiment 1, we suggested that children recognized that there is a link between objects' causal properties and their category membership. Here, we make a similar argument that 4-year-olds recognize that objects' insides are related to their category membership, whereas the nature of a sticker placed on the back of an object is independent of its category.

An alternative possibility is that children do not make a principled inference about the informants' knowledge of the category membership of the objects. Instead, they simply recognize that knowledge about insides indicates that the individuals are more familiar with the objects than knowledge about outsides. Birch and Bloom (2002), for example, suggested that preschoolers use familiarity to make inferences about individuals' proper names. Although this is a possibility, familiarity and knowledge of category membership seem confounded. In both conditions, the confederates are knowledgeable about a nonobvious property of the object that is revealed during the course of the experiment. If children believe that the confederates in the insides condition are more familiar with these objects because of the nature of the property, then they must recognize that the property is relevant to the objects' conceptual structure.

It is important to note that the inferences 3-year-olds made were different from those made by 4-year-olds. There are two possible reasons for this difference. One is that the younger children were unable to register the different expertise between the two informants. Although 3-year-olds struggled with the knowledge question compared to 4-year-olds, this difference was not significant. Even those 3-year-olds who responded with 100% accuracy on the knowledge questions performed at chance on the ask and endorse questions. These responses potentially reflect their not recognizing the relation between objects' internal or external parts and category membership. The 3-year-olds' data do suggest that children are not responding based on recognizing that one informant is more familiar with one set of objects; Birch and Bloom (2002) found that 2-year-olds could make inferences about familiarity. Future research should attempt to tease apart children's knowledge of conceptual structure from the simpler idea that children register that one individual is more familiar with the objects.

General Discussion

Several experiments have found that preschoolers use informants' past reliability when learning conventional or constructed information. We expand on these findings in two ways. First, children appear to use individuals' ability to predict objects' causal efficacy accurately to make subsequent inferences about a new object's label. By age 4, children can transfer their understanding of speakers' reliability about an object's properties to another inference regarding that object. This is consistent with investigations that have demonstrated that children make inferences about object's intended functions based on speakers' reliability about familiar objects' labels (Birch et al., 2008; Koenig & Harris, 2005).

Second, there appears to be a limit to the extent young children will transfer this knowledge. When shown that individuals were knowledgeable about different internal properties of objects, 4-year-olds used that expertise to make inferences about the label of a new object with the same insides. When shown individuals with expertise about a nonobvious external property, 4-year-olds failed to use that information systematically to make the same inference. Younger children did not use either piece of information. These data are consistent with previous research suggesting that children do not relate object's insides with their category membership or causal properties until age 4 (Gottfried & Gelman, 2005; Sobel et al., 2007). The results from Experiment 2 suggest that 4-year-olds use informants' expertise only when it is relevant to the inference they are asked to make.

These data are also consistent with the hypothesis that 4-year-olds are simply better at monitoring multiple sources of information, similar to the way in which 3-year-olds struggle to understand cases where both informants have probabilistic levels of accuracy (i.e., neither informant is always right or always wrong). Pasquini et al. (2007) and other investigations that consider 3-year-olds' inferences given uncertain data (e.g., Corriveau et al., 2009) interpret their findings as 3-year-olds mistrusting informants who have made any error, regardless of the comparative frequency of those errors. Although children never observe an informant err in these experiments, the younger children might be more likely to treat the experts' ignorance as an incorrect response, given that there is no a priori reason for one informant to only know about one kind of causal property.

Future research should attempt to determine whether the developmental difference observed in Experiment 2 was a result of 3- and 4-year-olds' emerging information-processing abilities or their ability to understand that the informants' expertise is related to the inference(s) they are asked to make. One possibility would be to introduce 3- and 4-year-olds to pictures of familiar stimuli from two distinct classes (e.g., animals and artifacts) and two confederates. One confederate would accurately label all of the animals and be ignorant about the artifacts' labels; the other would show the reverse knowledge. If both 3- and 4-year-olds clearly distinguish animals from artifacts (e.g., as suggested by Gelman, 2003; Rosengren, Gelman, Kalish, & McCormick, 1991, among others), both age groups should have little difficulty when asked who could provide labels for pictures of novel animals and artifacts. This would suggest that such information-processing difficulties are not a likely explanation. In contrast, if a similar developmental difference to Experiment 2 is found, the development of these information-processing abilities is a more likely explanation.

Another question that emerges from these experiments is the exact expectations children have about the relations between objects' causal properties, insides, and category membership. For example, most studies of the relation between insides and category membership present children with a conflict between the object's insides and perceptual similarity (e.g., Gelman & Wellman, 1991; Sobel et al., 2007). In these cases, 3-year-olds use perceptual similarity as a guide to category membership, whereas 4-year-olds use the object's insides. We do not believe that 3-year-olds have no understanding of the relation between object's insides and category membership, but for this object set (simple artifacts that were similar to the ones used by Sobel et al., 2007), 3-year-olds might have less insight into the relation between insides and labels. An open question is whether 3-year-olds would respond differently if the artifacts were made more complicated so that the relevance of their insides was clearer. If preschoolers are just developing the ability to monitor informants' expertise, then a manipulation like this should not matter. If, instead, they are integrating that ability with their understanding of relevance, this might improve 3-year-olds' use of the informants' expertise.

One conclusion we can draw from these data is that by the time children are 4, they possess some of the cognitive abilities to distinguish reliable informants among different domains of knowledge. This might be the basis of the more metacognitive ''division of cognitive labor'' findings by Keil and colleagues in which children recognize that individuals have access to different pieces of knowledge, related to domains in which they are shown to be experts. This is consistent with the fact that 4-year-olds in both experiments showed at least some differences in the distributions of responses to the ask and endorse questions. Deciding what the object is called (the endorse question) requires the child to choose which of the confederates generated the accurate label. In contrast, deciding who knows about the label (the ask question) potentially requires the child to choose which confederate would generate the accurate label, which might require better access to the process by which the child would make this decision.

To conclude, there is little doubt that young children can rely on other people as sources of causal knowledge. Children clearly learn language (or at least word referents), science, religious concepts, and the nature of extraordinary beings by relying on the information provided by others. However, how this is accomplished is not well understood. Harris and Koenig (2006) point out that ''children might . . . be equipped with heuristics or strategies for assessing the quality or plausibility of the testimony that is made available to them'' (p. 519). The present data suggest that children's causal knowledge is part of the ''heuristics and strategies'' that Harris and Koenig suggest are critical to learning from testimony. By the age of 4, children are not just using the most reliable source of information but rather are using their own causal knowledge to judge whether that reliability is relevant.

References

Birch, S. A., & Bloom, P. (2002). Preschoolers are sensitive to speaker's knowledge when learning proper names. *Child Development*, *73*, 434–444.

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three- and 4-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034.

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

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

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*, 948–957.

Clément, F., Koenig, M., & Harris, P. L. (2004). The ontogenesis of trust in testimony. *Mind and Language*, *19*, 360–379.

Corriveau, K. H., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*, 188–193.

Corriveau, K. H., Harris, P. L., Meins, E., Ferneyhough, C., Arnott, B., Elliot, L., et al. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, *80*, 750–761.

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology*, *27*, 331–342.

Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist? Developmental shifts in ways of clustering knowledge. *Child Development*, *75*, 918–931.

de Saussure, F. (1966). *Course in general linguistics*. New York: McGraw-Hill.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

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

Gelman, S. A., & Gottfried, G. M. (1996). Children's causal explanations for animate and inanimate motion. *Child Development*, *67*, 1970–1987.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understanding of the non-obvious. *Cognition*, *38*, 213–244.

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

Gottfried, G. M., & Gelman, S. A. (2005). Developing domain-specific causal-explanatory frameworks: The role of insides and immanence. *Cognitive Development*, *20*, 137–158.

Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, *75*, 409–427.

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*, 505–524.

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science*, *9*, 76–96.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*, 757–758.

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the Division of Cognitive Labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*, 259–300.

Kemler Nelson, D. G., Almasy, L., Crowley, K., Duke, N., Gardner, J.A., Kiggins, V., et al. (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. (2000). Young children's use of function information to categorize artifacts: Three factors that matter. *Cognition*, *77*, 133–168.

Koenig, M., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *10*, 694–698.

Koenig, M., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277.

Koenig, M. A., & Harris, P. L. (2008). The basis of epistemic trust: Reliable testimony or reliable sources? *Episteme*, *4*, 264–284.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development*, *73*, 1073–1084.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, *4*, 389–396.

Nurmsoo, E., & Robinson, E. J. (2009). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science*, *12*, 41–47.

Pasquini, E., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226.

Rosengren, K., Gelman, S. A., Kalish, C., & McCormick, M. (1991). As time goes by: Children's early understanding of biological growth. *Child Development*, *62*, 1302–1320.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054–1070.

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

Sobel, D. M., Yoachim, C. M., Gopnik, A., Meltzoff, A. N., & Blumenthal, E. J. (2007). The blicket within: Preschoolers' inferences about insides and causes. *Journal of Cognition and Development*, *8*, 159–182.