

Naive theory and transfer of learning: When less is more and more is less

VLADIMIR M. SLOUTSKY and MARGIE A. SPINO
Ohio State University, Columbus, Ohio

In this research, we examine how young children perform induction in the domain of biology. According to the essentialist view, even young children have a priori knowledge that some predictors of biological properties are central, whereas others are peripheral. According to the flexible-learning view, the roles of predictors are not fixed, and young children flexibly learn the predictive value of each predictor. In the experiment reported, 4- to 5-year-olds and 11- to 12-year-old preadolescents were taught the predictive values of features that were (by the essentialist account) either central or peripheral for predicting the biological properties of animals. Retention of the learned predictors was then tested immediately after learning, 1 h after learning, and 2.5–3.5 months later. The results indicate that 2.5–3.5 months after learning, young children retained learned predictors well regardless of their putative centrality, whereas preadolescents were likely to retain and recall only central, but not peripheral, predictors. These results do not support the essentialist claim that young children differentiate between central (or essential) and peripheral predictors in the domain of biology and, thus, challenge the naive-theory view of induction in young children.

Inductive generalization is crucially important for the expanding of knowledge to novel situations: Upon learning that a particular bird has hollow bones, one would expect another bird also to have hollow bones. Although there is little disagreement that even young children can perform simple inductions, the specific processes underlying their inductions remain highly contested.

According to one theoretical position, from the beginning of conceptual development, induction is a function of theoretical beliefs about the world (Gelman & Coley, 1991; Keil, Smith, Simons, & Levin, 1998). Even young children are said to have organized knowledge (or naive theories) in some knowledge domains, such as biology, and they believe that some properties of biological organisms are more central, or essential, than others (Gelman & Coley, 1991). According to this view, the belief that some properties are essential is not learned but, rather, exists early in development as an a priori assumption that directs all learning, including induction (Keil et al., 1998). This *essentialist assumption* directs induction of even young children, because they are said to believe that essential properties are causally related to unobserved properties (whereas peripheral properties merely co-occur with unobserved properties) and children understand that the former are better predictors of unobserved properties than are the latter (Keil et al., 1998). For example, the essential property *dogness* com-

municated by the label *dog* is more central for the predicting of *insides* of the animal than is *dog-like* appearance (Gelman & Coley, 1991; Keil et al., 1998). According to Gelman and Coley, “Names are embodiment of our theories” (p. 190). Therefore, when two even differently looking animals are called *dogs*, children should believe that these animals have many unobserved properties in common, and this belief supports their induction. In short, according to the essentialist position, even young children believe that, in the domain of biology, linguistic labels are central, whereas perceptual similarity is peripheral, and that these *central* and *peripheral* roles are fixed.

However, there is evidence that reliance on labels in the course of induction exhibits greater flexibility than would be expected under the essentialist assumption of centrality. For example, under some conditions, children are more likely to induce properties on the basis of perceptual similarity than on the basis of matching labels: 3- and 4-year-olds are more likely to rely on similarity of motion than on the matching linguistic label (Mak & Vera, 1999). There is also evidence that young children consider a linguistic label as a feature contributing to the overall similarity of compared entities and, through similarity, to induction (Sloutsky & Lo, 1999; Sloutsky, Lo, & Fisher, 2001). If the values of different predictors change in the course of learning, it is likely that the importance of different predictors is learned rather than assumed a priori. Such a learning account would explain the flexibility often exhibited by young children (e.g., Smith, Jones, & Landau, 1996) better than would the essentialist position, which presumes fixed and nonflexible roles of central and peripheral features.

One way of distinguishing between these positions is to examine the ability of young children to acquire and retain

This research was supported by Grant BCS 0078945 from the National Science Foundation to V. M. S. We thank Anna Fisher, Uri Hasson, Valerie Kuhlmeier, Chris Robinson, and Aaron Yarlas for their helpful comments. Correspondence concerning the article should be addressed to V. M. Sloutsky, Center for Cognitive Science, Ohio State University, 208C Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210 (e-mail: sloutsky.1@osu.edu).

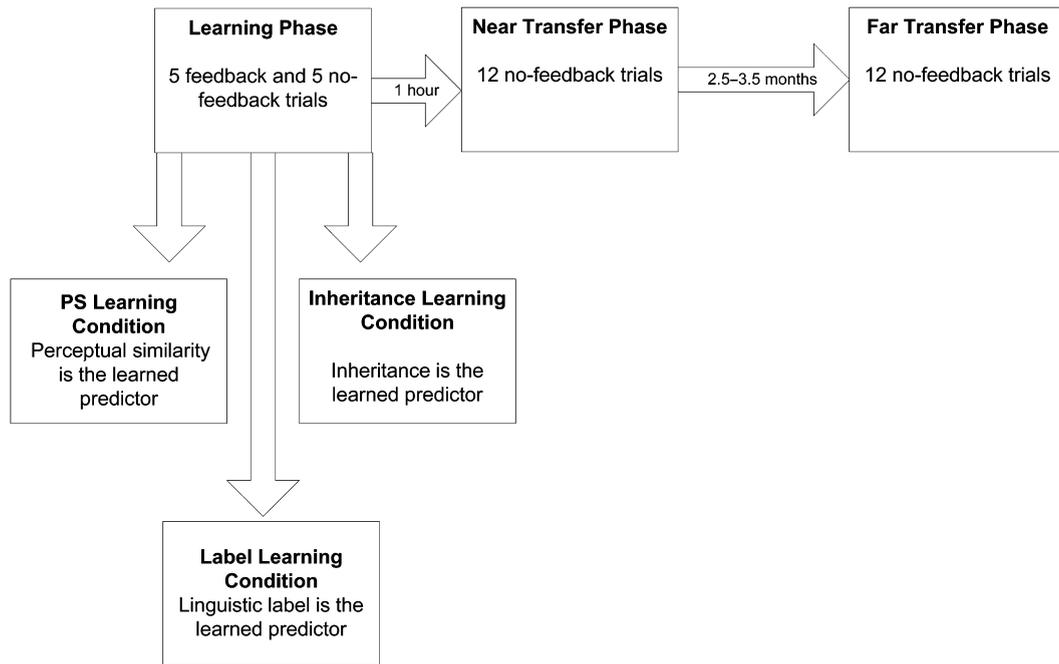


Figure 1. The overall design of the experiment.

knowledge that is inconsistent with their putative essentialist assumptions. It is well known that when people have organized knowledge (e.g., beliefs, theories, or schemas), this knowledge may affect their acquisition of new knowledge by facilitating acquisition of knowledge-consistent information and hindering acquisition of knowledge-inconsistent information (Alba & Hasher, 1983). For example, when children have gender stereotypes (stereotypes are a form of organized knowledge), they better retained information that fitted their stereotypes (Liben & Signorella, 1980).

Distortion, or knowledge inconsistency, effects occur when parts of knowledge-consistent information are absent from the encoding phase: Participants are likely to commit errors of commission by falsely recalling and recognizing absent information. For example, when presented with words related to the word *doctor* (e.g., *nurse*, *hospital*, *bed*, etc.), participants falsely recalled and recognized *doctor*, which had not been presented (see Roediger, 1996, for a review). Knowledge inconsistency effects also occur when incoming information is inconsistent with prior knowledge—in which case, participants are likely to commit errors of omission by filtering out information inconsistent with their beliefs and expectations (Stangor & McMillan, 1992).

To distinguish between the two approaches to young children's induction, we created a task in which young children and preadolescents were trained to induce biological properties of offspring by relying on a single predictor: common inheritance, common label, or perceptual resemblance. According to the essentialist position, for young children, appearance information is more peripheral in the domain of biology than is the common label. Therefore, if chil-

dren do assume differential centrality, they should better retain information I_1 honoring the assumed centrality (i.e., knowledge-consistent information), than information I_2 , violating the assumed centrality (i.e., knowledge-inconsistent information). At the same time, if the weights of predictors are learned rather than assumed, children should flexibly acquire and retain both I_1 and I_2 . Previous research (Lo & Sloutsky, 2001) has indicated that preadolescents do have theoretical knowledge in the domain of biology (i.e., knowledge that shared inheritance is a more reliable predictor of common biological properties than is appearance), and we expected them to exhibit knowledge inconsistency effects—poorer retention of theory-inconsistent information. At the same time, we expected young children to exhibit flexible learning and, thus, accurate retention of putatively central and peripheral predictors. The latter prediction is in contrast with the predictions of the essentialist assumption, which would expect the same theory inconsistency effects in both age groups.

METHOD

Participants

The participants were 78 young children (mean age = 4.98 years, $SD = 0.31$ years; 33 boys and 45 girls) and 45 preadolescents (mean age = 12.06 years, $SD = 0.32$ years; 21 boys and 24 girls), all recruited from schools located in upper middle class suburbs of Columbus, Ohio.

Design, Materials, and Procedure

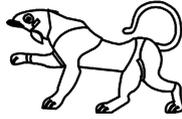
The general task was to induce a biological property from one of several test animals to a target animal. Each test animal was perceptually similar to the target, had a common inheritance, or shared a linguistic label with the target, and the participants could rely on each

A. Learning Trial



Target: A guga

Question: *Does this guga have pink, white, or blue fat?*



Test 1: A pofi who gave birth to the target.
Has pink fat

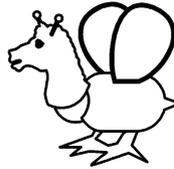


Test 2: A guga who played with the target.
Has white fat



Test 3: A boto who played with the target.
Has blue fat

B. Near Transfer Trial (PS vs. LAB + INH).

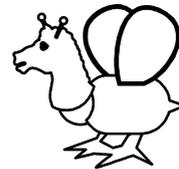


Target: A lolo.

Question: *Does this lolo have rough skin or soft skin?*



Test 1: A lolo who gave birth to the target.
Has rough skin



Test 2: A jaja who played with the target.
Has soft skin

Figure 2. Examples of trials in the (A) learning and (B) near transfer phases.

of these predictors in their induction. The experiment had a mixed design, with age and learning condition as between-subjects factors and experimental phase as a within-subjects factor (see Figure 1, depicting the overall design). There were three learning conditions, and in each condition the participants were trained to rely on a single predictor when inducing biological properties: perceptual similarity (PS), shared inheritance (INH), or shared label (LAB). The participants were randomly assigned to one of the learning conditions. There were three phases: learning, near transfer, and far transfer. A female experimenter administered the former two phases on a Dell Inspiron 3500 laptop computer. A male experimenter administered the latter phase manually.

Learning phase. The learning phase consisted of 10 trials. On each trial, four pictures of animals (a target and three test animals) were presented to the participant, with the Test 3 animal being visually identical to the target (see Figure 2A for examples of the pictures). The pictures were line drawings, each measuring approximately 10×7.5 cm, created so as not to resemble any actually

existing animals. The pictures were presented on a computer screen. Then labels were introduced, so that only Test 2 shared the label with the target. The labels were two-syllable artificial count nouns (e.g., *a guga*). Finally, the participants were told that Test 1 shared inheritance with the target (it was said to “give birth” to the target). The location and order of each test animal, relative to the target, was counterbalanced across trials. Biological properties were attributes referring to the insides of animals; they were randomly assigned to the test animals. The list of labels and biological properties is presented in the Appendix.

In short, each of the test animals shared only one predictor with the target animal: One shared a linguistic label, one was perceptually similar, and one shared inheritance (see Figure 2A for an example of a learning trial). Although “giving birth” may appear too complex for young children, our previous research had indicated that young children did not have difficulty with this marker of inheritance (Lo & Sloutsky, 2001). In addition, differential difficulty of predictors for 4- to 5-year-olds was of no concern, because we ex-

C. Far Transfer trial (INH vs. LAB + PS).

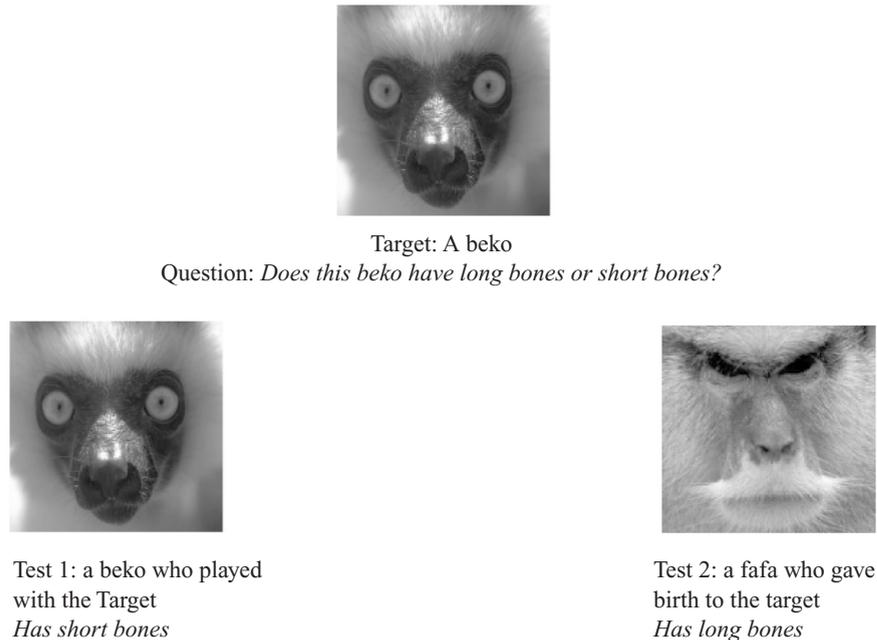


Figure 2 (Continued). Example of trials in the (C) far transfer phase.

pected (and indeed found) high degrees of learning across predictor types. Finally, each of the test animals was described as having an unobserved biological property (e.g., *white fat vs. blue fat vs. pink fat*), and the participants were asked to decide which of the three properties the target had. Detailed instructions are presented in the Appendix.

The goal of learning was to train the participants to perform induction by consistently relying upon a single predictor—PS, INH, or LAB—and the participants were trained by means of feedback. On each trial, a participant was asked to induce a property. If he or she relied on a condition-consistent property (e.g., on PS in the PS learning condition), *That's right* feedback was given to him or her; otherwise, *Try again* feedback was given. In the latter case, the incorrect choice was removed from the display, and the participant was asked to choose again from the remaining two test animals. After making his or her choice, the participant was again provided with feedback. If the choice was condition consistent, *That's right* feedback was provided, and the participant moved to the next trial; otherwise *Try again* feedback was given. In this case, the incorrect choice was removed from the display, and the task was repeated with the remaining test animal and the target animal. Feedback was provided only on the first five learning trials; the last five learning trials had no feedback. Regardless of the correctness of the choice, on each no-feedback learning trial, the participant was allowed only one induction choice.

The participants proceeded to the transfer phases only if they made at least four out of five learned responses on the five no-feedback learning trials. Ten 4- to 5-year-olds (5 in the INH condition and 5 in LAB condition) did not meet this criterion, and 13 4- to 5-year-olds were not available at the time of near or far transfer. These participants, therefore, were included only in the analysis of the learning phase. All other participants moved to the transfer phases.

Near transfer phase. Near transfer (which took place approximately 1 h after learning) had the same induction task as the learning phase, except that no feedback was provided for any trials. The

task consisted of 12 trials. The materials were identical to those in the learning phase, except that two, rather than three, test animals were presented, with Test 1 sharing two of the three predictors with the target, and Test 2 sharing the remaining predictor. Each possible pairing of predictors (i.e., PS vs. INH + LAB, INH vs. PS + LAB, and LAB vs. PS + INH) appeared four times (see Figure 2B for an example of a near transfer trial). This change was needed in order to have a conservative measure of learning by making the transfer task different from the learning task.

Far transfer phase. The far transfer phase took place 2.5–3.5 months after learning ($M = 99.21$ days and $SD = 47.67$ days for the 5-year-olds; $M = 71.08$ days and $SD = 13.98$ days for the 12-year-olds). The task was identical to that in the near transfer phase. The procedure was similar to that in the learning and near transfer phases, with the following exceptions: (1) Photographs of unusual animals, rather than line drawing pictures, were used; (2) stimuli were administered manually, rather than by a computer program; and (3) a male, rather than a female, experimenter administered the far transfer phase (see Figure 2C for an example of a far transfer trial). The photographs depicted unusual animals (see Figure 2C), each photograph measuring 8.75×8.75 cm. These changes were necessary to eliminate the possibility of researcher, computer program, or stimuli, serving as event “tags” that could remind the participants of predictors learned during the learning phase. At the end of the far transfer phase, the participants were asked whether they remembered the learning experiment, and if they did, they were asked to recall the predictor of the properties in the learning phase. They were then debriefed and were told that, in real life, shared inheritance is a better predictor of unobserved biological properties than are other predictors.

RESULTS

In the learning phase, the majority of the participants successfully learned condition-consistent predictors. Pro-

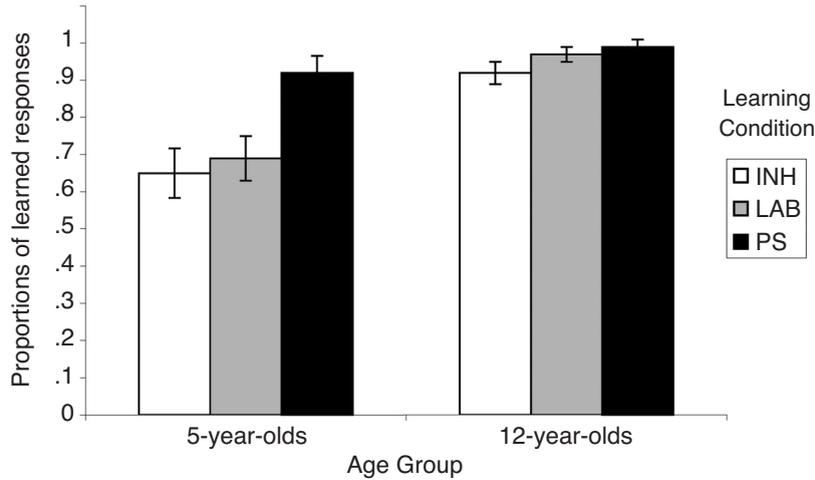


Figure 3. Proportions of learned responses by age and learning condition in no-feedback trials of the learning phase. Error bars represent standard errors of the mean. INH, shared inheritance; LAB, shared label; PS, perceptual similarity.

portions of learned responses across five no-feedback learning trials are presented in Figure 3. These data were subjected to a 3 (learning condition) \times 2 (age group) between-subjects analysis of variance (ANOVA). There was a significant main effect of learning condition, with a higher proportion of learned responses in the PS condition (92%) than in the INH condition [73%; $F(2,117) = 3.49$, $MS_e = 0.08$, $p < .05$; post hoc Tukey test, $p < .01$], and a significant main effect of age, with the preadolescents exhibiting a greater proportion of learned responses (92%) than the 4- to 5-year-olds [75%; $F(1,117) = 11.39$, $MS_e = 0.08$, $p < .002$]. At the same time, the interaction was not significant [$F(2,117) = 2.4$, $p > .09$].

These findings indicate that although the majority of the participants in both age groups learned successfully,

(1) the preadolescents exhibited better learning than did the 4- to 5-year-olds, and (2) the INH condition was somewhat more difficult than the PS condition. None of these findings is surprising, given the age differences and the fact that, unlike perceptual similarity or label, shared inheritance is an abstract nonperceptible property.

The results of the near transfer phase are presented in Figure 4. Note that each predictor was pitted against a pair of remaining predictors, thus yielding three combinations: PS versus INH + LAB, INH versus PS + LAB, and LAB versus PS + INH, with each combination presented four times. If the participants randomly selected a single predictor, they should have selected the learned predictor in 33% of the responses; if they randomly selected a pair versus a single predictor, they should have selected the learned

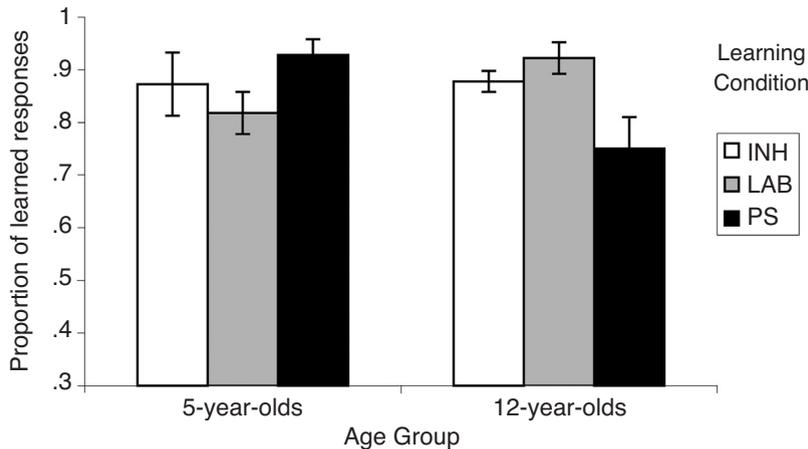


Figure 4. Proportions of learned responses by age and learning condition in the near transfer phase. Error bars represent standard errors of the mean. INH, shared inheritance; LAB, shared label; PS, perceptual similarity.

predictor in 66% of the responses, whereas if they were merely guessing, they should have selected the learned predictor in 50% of the responses. Across all conditions, the participants selected the learned predictor reliably above 66% (all one-sample t s > 3.4 , p s $< .001$), except for the preadolescents in the PS condition, who were no different from 66% ($t < 1$).

The proportions presented in Figure 4 were subjected to a 3 (learning condition) \times 2 (age group) between-subjects ANOVA. Neither main effect was significant (both F s < 1), whereas there was a significant age \times condition interaction [$F(2,94) = 4.4$, $MS_e = 0.05$, $p < .02$]. The analysis of the interaction indicated that whereas the preadolescents were less likely to retain the learned predictor in the PS condition than in the other two conditions [$F(2,43) = 3.7$, $MS_e = 0.06$, $p < .02$; post hoc Tukey tests, p s $< .05$], 4- to 5-year-olds were more likely to retain the learned predictor in the PS condition than in the LAB condition [$F(2,53) = 3.2$, $MS_e = 0.03$, $p < .05$; post hoc Tukey test, $p = .05$]. Although the 5-year-olds' results are not surprising, because the PS condition may have lower task demands than the LAB condition (i.e., unlike the identity of label, PS does not have to be remembered), the results for the preadolescents are quite informative. These results indicate that the PS predictor that has low task demands is the least likely to be retained even shortly after learning.

The results of far transfer are presented in Figure 5. Young children were reliably above 66% in the PS and INH conditions (one-sample t s > 3.5 , p s $< .01$), and they were marginally above 66% in the LAB condition ($p = .08$). At the same time, the preadolescents were significantly above 66% in the INH condition ($t > 24$, $p < .0001$), not different from 66% in the LAB condition, and not different from 33% in the PS condition (both t s < 1).

The proportions presented in Figure 5 were subjected to a 3 (learning condition) \times 2 (age group) between-subjects

ANOVA. There was a significant main effect of age, with the 4- to 5-year-olds exhibiting greater proportions of condition-consistent responses than did the preadolescents [78% vs. 64%; $F(1,94) = 13.3$, $MS_e = 0.03$, $p < .0001$]. There was also a significant main effect of the learning condition, with greater proportions of condition-consistent responses in the INH condition (90%) than in the PS or the LAB condition [60% and 66%, respectively; $F(2,94) = 30.38$, $MS_e = 0.03$, $p < .0001$]. The latter main effect was largely driven by a significant interaction [$F(2,94) = 20.77$, $MS_e = 0.03$, $p < .0001$], indicating that whereas for the 4- to 5-year-olds the probability of retaining the predictor did not differ significantly across the conditions, the preadolescents were most likely to retain the predictor in the INH condition (96%), less likely in the LAB condition (63%), and least likely in the PS condition (35%; post hoc Tukey tests, all p s $< .0001$). Furthermore, the preadolescents in the PS condition were significantly less likely than the 4- to 5-year-olds to retain the learned predictor [$t(34) = 8.78$, $p < .0001$], whereas there were no significant differences in the other two conditions.

After completing the far transfer phase, the participants were asked whether they remembered participating in the learning phase; all the 4- to 5-year-olds and all but 2 of the preadolescents remembered it. However, when asked to explicitly recall the learned predictor, the majority of the 4- to 5-year-olds in each condition could not verbalize the predictor: Only 29% of the 4- to 5-year-olds in the INH condition and 33% in the PS condition recalled the learned predictor. There were also 8 4- to 5-year-olds (14.8%) who incorrectly recalled the learned predictor (3 incorrectly recalling INH, 3 incorrectly recalling LAB, and 2 incorrectly recalling PS), whereas the remaining children did not recall any predictor. These relatively low rates of recall of the learned predictors, taken in conjunction with high rates of transfer, suggest that the 4- to 5-year-olds simply

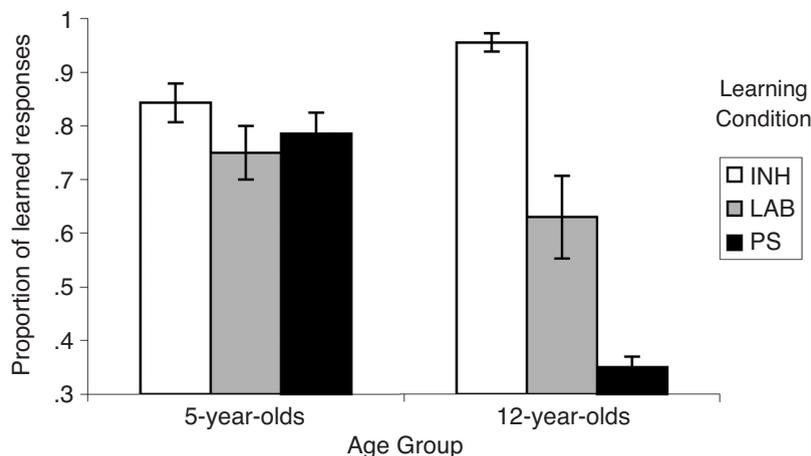


Figure 5. Proportions of learned responses by age and learning condition in the far transfer phase. Error bars represent standard errors of the mean. INH, shared inheritance; LAB, shared label; PS, perceptual similarity.

had difficulties verbalizing, but not retaining, the predictors. At the same time, although every preadolescent in the INH condition correctly recalled the learned predictor, nobody correctly recalled the predictor in the PS condition, and only 5 preadolescents (33%) correctly recalled the predictor in the LAB condition. Importantly, the 12-year-olds in the PS and LAB conditions did not merely fail to state the learned predictor; rather, they false alarmed on INH. In the PS condition, 60% erroneously recalled INH as the learned predictor, 20% erroneously recalled LAB as the learned predictor, and another 20% did not recall any predictor. In the LAB condition, 60% incorrectly recalled INH and 33% correctly recalled LAB as the learned predictor. Note that there was no single correct recall or false alarm on PS as a predictor. In short, the 4- to 5-year-olds were likely to accurately retain all predictors, whereas the preadolescents accurately retained INH and (to a lesser degree) LAB, but not PS, thus demonstrating knowledge inconsistency memory effects.

DISCUSSION

Overall, the majority of the 4- to 5-year-olds and the preadolescents learned successfully to rely on different predictors, with the preadolescents exhibiting somewhat higher acquisition rates than did the young children. An hour after learning, the preadolescents, but not the young children, exhibited lower near transfer for PS as a predictor. After a 2.5- to 3.5-month delay, the preadolescents exhibited low transfer and markedly decreased memory for PS, whereas the 4- to 5-year-olds accurately retained and transferred the learned predictor in all the learning conditions. The preadolescents also exhibited a high level of false recall in the PS condition, falsely recalling PS as either the INH or the LAB predictor. In short, as was predicted, the preadolescents (who are more knowledgeable than young children) exhibited knowledge inconsistency effects—poorer retention of knowledge-inconsistent information. At the same time, the young children exhibited flexibility in learning, accurately retaining putatively central and peripheral predictors.

The findings reported offer new evidence about the processes underlying young children's induction. The results do not support the essentialist claim of feature centrality: Young children accurately retained those predictors that were claimed to be theoretically central (i.e., linguistic labels), those that were claimed to be theoretically peripheral (i.e., perceptual similarity), and those that may be unfamiliar (i.e., inheritance information). At the same time, preadolescents (who could be considered a baseline group for the possession of organized knowledge) exhibited effects of organized knowledge: They retained and recalled theoretically important predictors (e.g., inheritance information and, to a lesser extent, labels) better than theoretically peripheral predictors (i.e., perceptual similarity). Furthermore, they falsely recalled theoretically important features being predictors during the learning phase, and they did not recall theoretically peripheral

predictors. The fact that young children accurately retained putatively central and peripheral predictors supports a flexible-learning account of predictor weights, rather than the essentialist account, which assumes fixed importance of different predictors.

Young children's accurate retention of different predictors in the transfer phase is unlikely to be a function of an *event memory*, because the far transfer phase differed from the learning phase in materials, equipment, and experimenter. In addition, because the participants in both age groups remembered participating in the learning experiment equally well, it seems unlikely that results stem from the learning phase being a more memorable event for 4- to 5-year-olds than for preadolescents. Therefore, it is unlikely that higher transfer rates in young children stem from a higher salience of the specific experimental event.

Could these results have stemmed from the fact that the task tapped into implicit, rather than explicit, processes (there were no instructions to learn or remember predictors in the learning phase and no reminding of the learning phase during the far transfer phase), with young children being more fluent in implicit processes than are preadolescents? Because implicit processes do not decrease with age (Buchner & Wippich, 1998), it seems unlikely that the results stemmed from a higher fluency of young children in implicit processes. Furthermore, we measured both implicit processes (i.e., performance on the task in the far transfer phase) and explicit processes (i.e., recall after the far transfer phase). Preadolescents exhibited low performance on both measures in the PS condition, whereas they exhibited high performance on both measures in the INH and LAB conditions. At the same time, in all three conditions, young children exhibited high performance on the measure of implicit processes and low performance on the measure of explicit processes. Therefore, the reported differences seem to reflect knowledge inconsistency effects in preadolescents and flexible learning in young children, rather than differential fluency in implicit processes.

Overall, the presence of the knowledge inconsistency effect in preadolescents and accurate retention of different predictors by young children cast doubt on the essentialist claim that young children distinguish between central and peripheral predictors, suggesting instead flexible learning of the importance of the different predictors by young children.

REFERENCES

- ALBA, J. W., & HASHER, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*, 203-231.
- BUCHNER, A., & WIPPICH, W. (1998). Differences and commonalities between implicit learning and implicit memory. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 3-46). Thousand Oaks, CA: Sage.
- 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* (pp. 146-196). New York: Cambridge University Press.
- KEIL, F. C., SMITH, W. C., SIMONS, D. J., & LEVIN, D. T. (1998). Two dogmas of conceptual empiricism: Implications for hybrid models of the structure of knowledge. *Cognition*, *65*, 103-135.

LIBEN, L. S., & SIGMORELLA, M. L. (1980). Gender-related schemata and constructive memory in children. *Child Development*, **51**, 11-18.

LO, Y.-F., & SLOUTSKY, V. M. (2001). Effects of multiple sources of information on induction in young children. In J. Moore & K. Stenning (Eds.), *Proceedings of the XXIII Annual Conference of the Cognitive Science Society* (pp. 564-569). Mahwah, NJ: Erlbaum.

MAK, B. S. K., & VERA, A. H. (1999). The role of motion in children's categorization of objects. *Cognition*, **71**, B11-B21.

ROEDIGER, H. L., III (1996). Memory illusions. *Journal of Memory & Language*, **35**, 76-100.

SLOUTSKY, V. M., & LO, Y.-F. (1999). How much does a shared name make things similar? Pt. 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology*, **6**, 1478-1492.

SLOUTSKY, V. M., LO, Y.-F., & 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.

STANGOR, C., & MCMILLAN, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin*, **111**, 42-61.

APPENDIX

Biological Properties Used in Learning and Transfer Phases

Learning Phase	Transfer Phases
(1) Blue blood, (2) green blood, (3) yellow blood	(1) Thick blood, (2) thin blood
(1) Yellow bones, (2) red bones, (3) white bones	(1) Short bones, (2) long bones
(1) Red bubbles, (2) blue bubbles, (3) white bubbles	(1) Small bubbles, (2) large bubbles
(1) White heart, (2) green heart, (3) pink heart	(1) Wide heart, (2) narrow heart
(1) Brown brain, (2) green brain, (3) red brain	(1) Large brain, (2) small brain
(1) Red muscles, (2) blue muscles, (3) yellow muscles	(1) Short muscles, (2) long muscles
(1) Blue fat, (2) white fat, (3) pink fat	(1) Soft fat, (2) hard fat
(1) White tongue, (2) blue tongue, (3) green tongue	(1) Long tongue, (2) short tongue
(1) White stomach, (2) green stomach, (3) blue stomach	(1) Small stomach, (2) large stomach
(1) Pink skin, (2) yellow skin, (3) white skin	(1) Rough skin, (2) soft skin
	(1) Small wrinkles, (2) large wrinkles
	(1) Can see at night, (2) cannot see at night
	(1) Has sugar inside the body, (2) has salt inside the body
	(1) Has thin hair inside its ears, (2) has thick hair inside its ears

Instructions

The important part of the instructions was: *I am going to show you pictures of funny-looking animals, tell you something about them, and then ask you some questions about these animals. Are you ready to start? Please, look at them. This is a guga (points to the target). This is another guga (points to Test 2), this is a pofi (points to Test 1), and this is a boto (points to Test 3, which looks exactly like the target). This one (points to Test 1) gave birth to this one (points to the target), this one (points to Test 2) and this one (points to Test 3) played a lot with this one (points to the target). The participants were asked to repeat labels and inheritance information, and label and inheritance information was repeated one more time, if a participant made at least one mistake. This pofi (points to Test 1) has yellow bones inside his body. This guga (points to Test 2) has red bones inside his body. And this boto (points to Test 3) has white bones inside its body. What does this guga (points to the target) have inside his body? Does he have red bones, yellow bones or white bones? After getting the participants' response, feedback was provided. In the PS learning condition, the participants should generalize on the basis of appearance, and therefore, white bones was the correct response; in the INH condition, they should generalize on the basis of common inheritance, and red bones was the correct response; and in the LAB condition, they should generalize on the basis of common labels, and yellow bones was the correct response.*