



Contents lists available at ScienceDirect

# Journal of Experimental Child Psychology

journal homepage: [www.elsevier.com/locate/jecp](http://www.elsevier.com/locate/jecp)



## The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes

Maria Fusaro\*, Kathleen H. Corriveau, Paul L. Harris

Harvard Graduate School of Education, Cambridge, MA 02138, USA

### ARTICLE INFO

#### Article history:

Received 14 September 2010

Revised 17 June 2011

Available online 29 July 2011

#### Keywords:

Selective trust

Testimony

Attribution

Halo effect

Traits

Early childhood

### ABSTRACT

Much recent evidence shows that preschoolers are sensitive to the accuracy of an informant. Faced with two informants, one of whom names familiar objects accurately and the other inaccurately, preschoolers subsequently prefer to learn the names and functions of unfamiliar objects from the more accurate informant. This study examined the inference process underlying this preference. We asked whether preschoolers make narrow inferences about informants, broader trait-based inferences, or more global evaluative inferences. We further asked what inferences preschoolers make about a potential informant based on distinctions in the unrelated domain of physical strength. The results indicate that preschoolers make relatively narrow inferences when observing individual differences in accuracy even though they are prone to global evaluative inferences when observing individual differences in strength. Preschoolers' burgeoning understanding of others as expert language users may underlie their selective endorsement of a more accurate informant.

© 2011 Elsevier Inc. All rights reserved.

### Introduction

For children to engage effectively in communication, they need to distinguish between trustworthy information sources and those that are prone to error. It is commonly assumed that preschoolers are gullible, believing everything they hear regardless of the source of the information. However, several researchers have recently shown that, in experimental settings, young children are more likely to accept claims from a previously accurate informant as compared with a previously errant informant (Birch, Vauthier, & Bloom, 2008; Jaswal & Neely, 2006; Koenig, Clément, & Harris, 2004; Koenig &

\* Corresponding author. Present address: UC Davis MIND Institute, Sacramento, CA 95817, USA.

E-mail address: [mfusaro@ucdavis.edu](mailto:mfusaro@ucdavis.edu) (M. Fusaro).

Harris, 2005). In these various studies, preschoolers are presented with two informants. During a familiarization phase, one informant correctly labels familiar objects and another informant incorrectly labels the same objects (e.g., calls a *shoe* a *duck*). During a subsequent test phase, the two informants label unfamiliar objects with different novel labels (e.g., *fep*, *mib*). Preschoolers are invited to say what they think the object is called. By 3 or 4 years of age, children selectively endorse the label offered by the previously accurate informant. They also identify him or her explicitly as the better of the two at labeling objects. Although these findings have been replicated in several studies, the nature of the differentiation that children make between the two informants is not yet clear. The current study aimed to shed light on the types of inferences children make when distinguishing between informants.

Children might make one of several types of inference. One possibility is that they make trait-like inferences about an informant's typical behavior. From exposure to a speaker's repeatedly inaccurate claims, children might infer that the speaker is prone to error, not very knowledgeable about the objects, or deceitful and unwilling to offer good information. Trait concepts both describe a series of actions over time and explain the underlying causes of human actions (Yuill, 1993). If children distinguish between informants using trait-like inferences, they may be able to predict their behavior in subsequent interactions and respond to each one appropriately. Thus, a trait-like distinction might explain the selective preferences that children demonstrate when offered conflicting information.

However, the use of trait-like inferences is problematic for explaining selective trust because other evidence suggests that preschoolers have not yet developed a flexible understanding of traits. Kindergarteners rarely use trait-relevant words in open-ended descriptions of others (e.g., Benenson & Dweck, 1986; Livesley & Bromley, 1973), and 5-year-olds are also less systematic than 7- and 8-year-olds in predicting a story character's behavior based on a report of his or her relevant past behaviors (e.g., Rholes & Ruble, 1984; Rotenberg, 1980). Even though young children have some global trait words in their vocabulary, such as *nice* and *mean*, their use of trait inferences is presumed to be tenuous until approximately 8 years of age (Alvarez, Ruble, & Bolger, 2001; Yuill, 1993). Liu, Gelman, and Wellman (2007) proposed that young children demonstrate mixed performance on trait-based reasoning tasks because trait attribution is a two-component process. Based on their review of the relevant data, young children tend to succeed on tasks that require either behavior-to-trait inferences or trait-to-behavior predictions but tend to fall short when both components are needed, as in behavior-to-trait-to-behavior predictions. The selective trust paradigm appears to fall in this latter category. During the familiarization phase, informants are distinguished for children by their behaviors but not by trait-relevant words, yet children must interpret informants' behavior during the test phase.

If trait-based inferences do not underlie preschoolers' distinctions between competing informants, another explanation for their success in the selective trust paradigm is needed. An alternative lean interpretation is that children use an inference strategy based on behavior matching (Astington, 1991; Heyman & Gelman, 2000; Rholes & Ruble, 1984). That is, children may predict that informants' behavior will be consistent from one situation to another very similar situation. In this case, success would depend on similarities between the behaviors that index informants' reliability during the familiarization phase of an experiment, such as object labeling, and the subsequent behaviors that children are asked to predict during the test phase. However, a subset of selective trust studies has presented children with the reliability status of informants on one behavior and then tested their endorsement of information supplied by the informants on another behavior. For example, Koenig and Harris (2005) found that, based on two informants' prior labeling accuracy, 3- and 4-year-olds selectively endorsed the reliable informant when they each demonstrated functions for novel objects (see also Birch et al., 2008). Similarly, Sobel and Corriveau (2010, Experiment 1) presented 4-year-olds with one informant who knew about the causal properties of objects (e.g., knew which objects would "make the machine go") and one who did not. When asked which informant would know the name of a novel object, children selectively turned to the informant with more causal knowledge about objects. Finally, Rakoczy, Warneken, and Tomasello (2009) presented 4- and 5-year-olds with puppets who differed both in the correctness of their labeling of familiar objects and in their performance of a set of familiar acts (e.g., drawing successfully with a functioning pen versus drawing unsuccessfully with a malfunctioning pen). Children went on to endorse the reliable puppet when the two puppets offered competing rules for a novel game. Taken together, these various findings suggest that children

are not using a narrow behavior-matching strategy as the basis for their predictions of informants' future reliability.

A third possibility is that preschoolers distinguish between reliable and unreliable informants based on global evaluative judgments about their behavior. They may evaluate agents as generally good or bad based on their actions and, as a result of a "halo effect", predict that subsequent behavior will take on the same valence. Thus, children may reason through the two-informant paradigm by evaluating the accurate informant as better, in a global sense, than one who makes mistakes. In line with Werner's (1957) classic account of developmental change, the hypothesis that children's reasoning progresses from very general to specific has been applied to many areas of child development, including trait attribution processes (Alvarez et al., 2001; Livesley & Bromley, 1973). Alvarez and colleagues (2001) found age differences in the extent to which children's behavior predictions were associated with trait inferences versus their evaluative judgments of story characters. For example, 9- and 10-year-olds made predictions about a previously generous story character's subsequent behavior that were aligned with their rating of the extent to which that character was generous. By contrast, 5- and 6-year-olds' behavior predictions were more closely related to their ratings of the character's goodness (or badness).

Children's generalizations from a character's social-moral behavior provide further evidence of such global reasoning. Cain, Heyman, and Walker (1997) presented 4- and 5-year-olds with stories using two dolls interacting with peers on a playground. One doll engaged in two prosocial acts by helping to push a child on a swing and inviting a new peer to his birthday party. The other doll hit another child on the playground and called the new peer a name. The children predicted that the prosocial character would engage in additional prosocial acts (rather than antisocial acts). However, they also made valenced predictions extending beyond the social domain, namely that the nice doll, more so than the mean doll, would demonstrate intellectual and athletic skills.

A similar pattern of results was found in a study of children's predictions about their own classmates' competencies (Stipek & Daniels, 1990). Kindergarteners, and to some extent fifth graders, made positively valenced predictions about a "nice" peer in their class that extended beyond the social domain (e.g., sharing cookies). As compared with a peer who was "not nice", they reported that the nice peer would also be more successful at an academic task (e.g., answering questions about a book) as well as an athletic task (e.g., jumping over hurdles). A "smart" peer in their class was also judged to be a better performer than a peer who was "not very smart" across these three domains. By eighth grade, children made finer-grained distinctions between these peers, for example, reasoning that the nice and not nice peers would be distinct in prosocial behavior but equivalent in their jumping ability. Thus, on the basis of these various findings, it is plausible that young children infer that an accurate informant is generally better than an errant one.

In summary, we have proposed three different inference processes that preschoolers could use when making judgments in the selective trust paradigm. First, they could distinguish between informants using trait-like inferences. Second, they could make inferences based on close behavioral similarities between the familiarization and test phases, selectively choosing an informant only in instances where the behavior in each phase is evidently similar. Finally, preschoolers may be subject to a halo effect, making global judgments about informants' behavior.

In the current study, we used a modified version of the selective trust paradigm to identify the type of inference process that children use when faced with competing informants. Participants viewed a video clip of two puppets (informants) interacting with a third puppet (dog). To compare children's inferences about accuracy with their inferences about another potentially quite distinct domain, children were assigned to either the Accuracy or Strength condition. In the Accuracy condition, one puppet consistently labeled familiar items correctly and the other consistently made errors. In the Strength condition, one puppet demonstrated physical strength by lifting various boxes and baskets off the table, whereas the other puppet attempted but failed to lift the same items off the table. During the test phase for both conditions, participants were asked to assess which puppet had labeled a novel object correctly and which puppet had lifted a wooden block successfully. We then examined children's predictions of each puppet's competence in five behavioral domains, including behaviors within the competence domain and behaviors outside of it. Finally, children rated each puppet on three traits: "strong", "smart", and "nice".

Each of the proposed inference processes would lead to a different pattern of evaluations regarding the target puppets. If children use trait-based evaluations, we would expect their positive or negative evaluations of each puppet to extend only to relevant trait-related behaviors and trait labels (i.e., Accuracy condition: demonstrations of artifact knowledge and being smart; Strength condition: demonstrations of physical strength and being strong). Outside of these trait-relevant behaviors, children would not systematically distinguish between the target puppets. If children use a behavior-matching strategy, we would expect their positive or negative evaluations of each puppet to extend only to the behavior that most closely resembles the behavior presented during familiarization. Responses regarding other behaviors and all three traits would be unsystematic. Finally, if children make global evaluations, two predictions follow. First, children's positive or negative evaluations of each puppet would extend to all behavior and traits regardless of their trait relevance. Second, any observed behavior that preschoolers recognize as positively or negatively valenced would trigger selective endorsement of informants' claims. Thus, observed differences in the physical strength of potential informants should also lead to selective trust. Taken together, these tests should illuminate the specificity or breadth of young children's social evaluations by considering the potential transfer across domains of competence.

## Method

### Participants

Participants were 60 preschoolers (age range = 3 years 0 months (3;0)–5;3, 29 girls and 31 boys). During the familiarization phase, via video, 30 children in the Accuracy condition (age range = 3;0–5;1,  $M = 4;3$ ,  $SD = 6.6$  months) assessed accurate versus inaccurate puppets and 30 children in the Strength condition (age range = 3;1–5;3,  $M = 4;1$ ,  $SD = 7.7$  months) assessed physically strong versus weak puppets. Children in each condition were split into "older" and "younger" groups using median age values (Accuracy condition: younger age range = 3;0–4;1,  $M = 3;8$ ,  $SD = 4.5$  months; older age range = 4;2–5;0,  $M = 4;6$ ,  $SD = 3.0$  months; Strength condition: younger age range = 3;2–4;3,  $M = 3;9$ ,  $SD = 4.1$  months; older age range = 4;4–5;4,  $M = 4;10$ ,  $SD = 3.6$  months). An additional 12 participants ( $n = 7$  in Accuracy condition,  $n = 5$  in Strength condition) completed the procedure but failed the final memory question, so that their data were not included for analysis.<sup>1</sup> Participants were recruited from a children's exhibit in a science museum serving predominantly middle-class families. All children in the relevant age range were invited to participate, excluding those with inhibitory disabilities (e.g., deafness) and those who did not speak English. The majority of families who were invited to participate accepted. Reasons for not participating included time constraints, lack of interest by parent or child, and child's refusal to view the video or wear headphones.

### Procedure

Children were interviewed individually by one experimenter while a parent observed without participating. The procedure involved three stages, as described below.

### Familiarization trials

The experimenter presented either Accuracy or Strength familiarization trials to the child via video on a laptop computer. In each case, three puppets were visible on the screen. In the center was a dog puppet who posed questions to the two target puppets located to her left and right. The two target puppets both were female, differentiated primarily by their shirt colors of yellow and red.

<sup>1</sup> The key finding of condition-related differences in behavior and trait generalization remain significant when the children who failed the memory test are included in the relevant chi-square tests. To avoid further increasing the complexity of the Results section, we do not report on the data for this group.

*Accuracy familiarization trials.* During each of four familiarization trials, the dog asked for a known object's name ("Can you tell me what this is called?"). One puppet informant labeled the object accurately (e.g., "That's a car"), and the other labeled it inaccurately (e.g., "That's a spoon"). Across the four familiarization trials, one puppet consistently labeled four known objects accurately (car, shoe, bottle, and ball), whereas the other puppet labeled them inaccurately (spoon, duck, apple, and cup). The speaking order of the puppets alternated across trials.

*Strength familiarization trials.* In the Strength condition, each trial began with the dog asking the two puppets, in reference to a box or basket, "Can you lift this?" In succession, each puppet said, "I will lift it", and attempted to do so. One puppet successfully lifted each item (basket, cardboard box, paper carton, and wooden box), whereas the other puppet attempted but failed to lift each item. The puppets alternated in being first to attempt to lift each item.

Immediately after each of the four familiarization trials, children were asked control memory questions for that trial (e.g., "Which one called that a shoe [or duck]?", "Which one lifted [or did not lift] that off the table?"). If the child responded, "don't know", made an error, or did not attend to the clip in the first place, the clip was replayed and the prompts were repeated. The order of the response probes (correct/incorrect label and lifted/did not lift) alternated across trials. Finally, for each condition, the identity of the accurate or strong puppet (i.e., red or yellow) was counterbalanced across children.

Following the four familiarization trials, children in both conditions responded to three types of questions in a fixed order: (a) two test trials (i.e., a labeling trial and a lifting trial), (b) behavior prediction probes, and (c) trait rating probes. The ordering of items within each of the three question types was counterbalanced (for test trials) or randomized (for behavior predictions and trait ratings) across participants. We chose to keep the order of question types fixed for two related reasons. First, it is plausible that the explicit judgments about the informants that children were asked to make in the trait ratings might have biased children's willingness to accept informants' claims in the test trials and children's predictions about future behavior. Thus, trait probes were always presented last. Second, to keep the test trials (i.e., selective trust trials) free from any potential carryover effects from subsequent questioning, we always presented those trials first. Each of the three question types is described in more detail below.

### *Test trials*

For the label test trial, the experimenter presented children with a picture of a novel object (a black rubber toilet bulb) and confirmed that its name was not known. Children were then invited to watch a video in which the object was set in front of the dog puppet, who asked the target puppets, "Can you tell me what this is called?" The two puppets each offered a different novel label for the object (e.g., "That's a dax", "That's a wug"). The experimenter probed children's label acceptance by asking, for instance, "The one in the red shirt said it's a dax. The one in the yellow shirt said it's a wug. What do you think it's called, a dax or a wug?" Both verbal responses (e.g., "wug", "what the red girl said") and pointing responses were accepted.

Children in each condition were also asked a parallel test question about object lifting. The experimenter presented children with a video clip of a wooden block being lifted off the table by an unknown puppet. Only the puppet's arms were visible coming onto the screen from directly above the table, making the lifter's identity ambiguous. Immediately following the video clip, children were asked which puppet had lifted the item ("Who lifted it, the one in the red shirt or the one in the yellow shirt?"). The ordering of the response choices was counterbalanced among children in each condition.

### *Behavior predictions*

Immediately following the label and lifting trials, children were asked questions about the puppets' expected performance on five behaviors in a random order: labeling unfamiliar objects, lifting objects, knowing what types of food particular animals eat, throwing basketballs into a hoop, and sharing cookies with the dog. To introduce each behavior, children were shown a picture of three pertinent objects and then were asked three questions. For example, for the novel labeling behavior, children were shown a picture of three novel objects and then were asked whether each puppet would be

“good” or “not very good” at saying the names of the objects. Finally, they were asked which puppet would be “better” at saying their names. Three objects were used for each behavior to ensure that ratings were not tied to a particular object.

#### Trait ratings

Immediately following the behavior predictions, children were asked a trio of questions about each of three traits. Thus, for strength, children were asked whether each puppet was “strong” or “not very strong” and which puppet was “stronger”. An equivalent trio of questions was asked for “smart” and “nice”.

#### Memory question

As a final check that children remembered the familiarization, children were asked a single recall question. In the Accuracy condition, children were asked, “Which one of these girls called the shoe a duck, the girl in the yellow shirt or the girl in the red shirt?” In the Strength condition, children were asked, “Which one of these girls didn’t lift the box, the girl in the yellow shirt or the girl in the red shirt?” The order of the shirt colors was counterbalanced across participants. Only children who correctly answered this question were included in the analyses.

## Results

Each of the following analyses was conducted using two-tailed tests. No differences were detected in the overall pattern of responses by girls and boys. Thus, gender was not included in subsequent analyses.

#### Label and lifting test trials

To examine condition- and age-related differences in test trial performance, two logistic regression analyses were conducted. We first tested whether children were more likely to choose the novel object label offered by the positively valenced puppet than the one offered by the negatively valenced puppet. Condition and age group (older or younger preschoolers) were included as between-participants factors. Note that two older preschoolers in the Strength condition did not respond to the label question and, thus, were excluded from this analysis. Descriptively, in the Strength condition, 77% of the older preschoolers and 60% of the younger preschoolers endorsed the label offered by the stronger puppet. In the Accuracy condition, 80% of the older preschoolers and 67% of the younger preschoolers endorsed the label offered by the accurate puppet. The logistic regression analysis revealed no significant effect of either condition,  $\beta = 0.25$ ,  $SE = 0.59$ , Wald’s  $\chi^2 = 0.18$ , *ns*, or age,  $\beta = 0.75$ ,  $SE = 0.60$ , Wald’s  $\chi^2 = 1.60$ , *ns*. Adding the interaction of Condition  $\times$  Age Group did not improve the fit of the model. Thus, the likelihood of choosing the label offered by the positive puppet was similar across age groups and conditions.

We next tested whether children were more likely to identify the ambiguous lifter as the positively valenced puppet, rather than the negatively valenced puppet, with condition and age group again included as between-participants factors. Descriptively, in the Strength condition, 80% of the older preschoolers and 80% of the younger preschoolers endorsed the label offered by the stronger puppet. In the Accuracy condition, 27% of the older preschoolers and 47% of the younger preschoolers endorsed

**Table 1**

Percentages of children favoring the positively valenced puppet on test trials by experimental condition.

Condition	Test trial	N	Percentage	Binomial test <i>p</i> value
Strength	Label	28 <sup>a</sup>	67.9	.089
	Lift	30	80.0	.002
Accuracy	Label	30	73.3	.018
	Lift	30	36.7	.201

<sup>a</sup> In the Strength condition, two children did not provide a response on the label test trial.

the label offered by the accurate puppet. The logistic regression analysis revealed a significant effect of condition,  $\beta = 1.96$ ,  $SE = 0.60$ , Wald's  $\chi^2 = 10.70$ ,  $p = .001$ . The effect of age was nonsignificant,  $\beta = -0.51$ ,  $SE = 0.59$ , Wald's  $\chi^2 = 0.76$ , *ns*. Adding the interaction of Condition  $\times$  Age Group in a subsequent model did not improve the fit. Thus, the main effect of condition indicates that, controlling for age group, the odds of identifying the positive puppet as the ambiguous lifter were 7.11 times higher for children in the Strength condition than for children in the Accuracy condition.

These findings were confirmed through comparisons with chance. Table 1 (upper rows) displays the percentages of children in the Strength condition who favored the strong puppet in the label and lift test trials. Binomial tests were used to compare the observed values with chance. Most children endorsed the object label that the strong puppet offered, although this trend fell short of significance. Most children identified the stronger puppet as the puppet that lifted the block. Table 1 (lower rows) displays the percentages of children in the Accuracy condition who favored the accurate puppet in the test trials. Most children endorsed the accurate puppet's label, but only a minority identified the accurate puppet as the puppet that lifted the block.

In sum, children extrapolated from a distinction in physical strength to a corresponding distinction both in later lifting and (nonsignificantly) in labeling behavior. By contrast, their extrapolation from a distinction in labeling accuracy was narrower. They extrapolated to a corresponding distinction in later labeling behavior but not to a distinction in lifting behavior.

In follow-up analyses, we examined children's individual responses to both the label and lift questions to further assess whether children had a greater tendency to generalize in the Strength condition than in the Accuracy condition. Of the 30 children in each condition, 24 in the Strength condition and 22 in the Accuracy condition correctly answered the test question associated with their familiarization (i.e., correctly answered the lift question in the Strength condition or correctly answered the label question in the Accuracy condition). Among these accurate responders, a greater proportion in the Strength condition (70%) than in the Accuracy condition (41%) generalized,  $\chi^2(1) = 3.74$ ,  $p = .053$ , Cramer's  $V = .29$ . Thus, analysis of individual response patterns confirmed that children used a distinction in strength to selectively endorse an object label more so than they used a distinction in labeling accuracy to make a prediction in an ambiguous lifting event.

### Behavior predictions

To generate scores for each behavior prediction, children received 1 point for reporting that the positively valenced puppet would be "good" at a given behavior, 1 point for reporting that the negatively valenced puppet would be "not very good" at it, and 1 point for reporting that the positively valenced puppet would be "better" (maximum score = 3). Mean scores are shown in Fig. 1 by

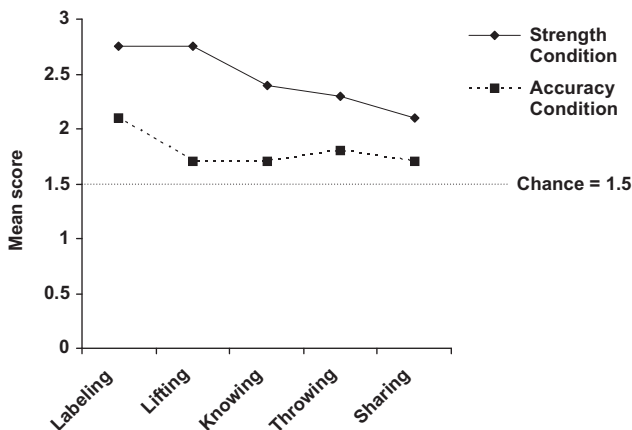


Fig. 1. Mean behavior prediction scores (maximum = 3) by condition (total  $N = 60$ ).

condition and behavior. We first used two-tailed *t* tests, comparing scores for each behavior with a 50% chance level of 1.5, to assess whether the positively valenced puppet was viewed as more competent than the negatively valenced puppet in each behavior domain. In the Strength condition, mean scores in all five behavioral domains were significantly greater than predicted by chance, with children favoring the stronger puppet over the weaker puppet ( $t_s > 3.7$ , all  $p_s \leq .001$ ). By contrast, in the Accuracy condition, mean scores exceeded chance levels only in the novel labeling prediction; the accurate puppet was seen as more competent than the inaccurate puppet only at labeling new objects,  $t(29) = 2.9$ ,  $p = .008$ , all other  $p_s > .10$ .

To confirm these findings, we conducted a three-way repeated-measures analysis of variance (ANOVA), with condition and age group as between-participants variables and behavior as a within-participants variable. In each condition, three children were excluded from this analysis due to missing one or more behavior scores (Strength: two younger and one older; Accuracy: one younger and two older). This analysis revealed a significant main effect of condition,  $F(1, 50) = 14.42$ ,  $p < .001$ ,  $d = 0.70$ ; mean scores were higher among children in the Strength condition ( $M = 2.48$ ,  $SD = 0.89$ ) than in the Accuracy condition ( $M = 1.80$ ,  $SD = 1.07$ ). A significant effect of behavior was also detected,  $F(4, 208) = 2.75$ ,  $p = .03$ . Furthermore, the Condition  $\times$  Behavior interaction neared significance,  $F(4, 208) = 2.13$ ,  $p = .079$ , suggesting that the difference between the two conditions varied by behavior.

Fig. 1 displays children's behavior predictions by Strength and Accuracy conditions. Inspection of the figure reveals that children in the Strength condition had higher behavior scores than children in the Accuracy condition, with this difference being greatest in the lifting task. To explore the Condition  $\times$  Behavior interaction further, the simple effect of condition was calculated for each of the five behavior types. The effect of condition was significant for labeling,  $F(1, 250) = 6.64$ ,  $p = .011$ ,  $d = 0.76$ , lifting,  $F(1, 250) = 18.40$ ,  $p < .001$ ,  $d = 1.42$ , and knowing about animals,  $F(1, 250) = 7.37$ ,  $p = .007$ ,  $d = 0.64$ , and was at trend levels for throwing basketballs into a hoop,  $F(1, 250) = 3.73$ ,  $p = .055$ ,  $d = 0.44$ , and sharing cookies,  $F(1, 250) = 2.95$ ,  $p = .087$ ,  $d = 0.48$ . In each case, scores were higher in the Strength condition than in the Accuracy condition.

In a follow-up analysis, individual children were categorized according to whether or not they made global behavior generalizations about the puppets. In the Strength condition, behavior prediction scores were summed, with the exception of scores for lifting (maximum = 12). In the Accuracy condition, behavior prediction scores were summed, with the exception of scores for labeling (maximum = 12). Based on a binomial test, for a given individual, the chance of observing 10 or more responses favoring the positive puppet, out of 12 trials, is less than 5% ( $p = .019$ , one-tailed). Thus, children with scores of 10, 11, or 12 were categorized as global generalizers, whereas children with scores below 10 were categorized as nongeneralizers. In each condition, three children were excluded due to missing one or more scores for behavior predictions. In total, 67% of children in the Strength condition ( $n = 11$  older preschoolers,  $n = 7$  younger preschoolers), but only 22% of children in the

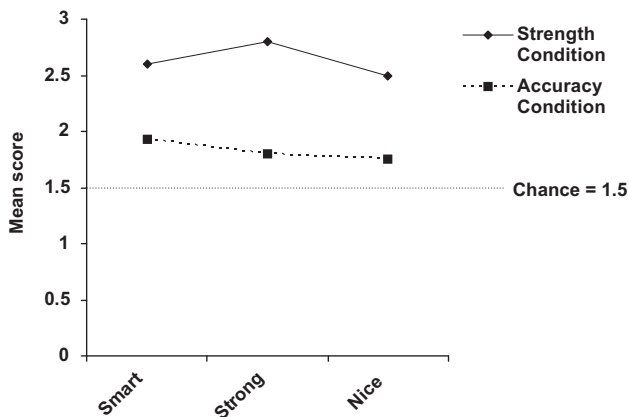


Fig. 2. Mean trait scores (maximum = 3) by condition (total  $N = 60$ ).



Accuracy condition ( $n = 4$  older preschoolers,  $n = 2$  younger preschoolers) were global generalizers. A chi-square test confirmed that a greater proportion of children in the Strength condition, as compared with the Accuracy condition, were global generalizers,  $\chi^2(1) = 10.80$ ,  $p = .001$ , Cramer's  $V = .45$ . These results confirm that children were more likely to make globally positive behavior predictions on the basis of physical strength than on the basis of accuracy.

### Trait ratings

On trait-rating questions, up to three points were similarly allocated with respect to each trait. For example, for “nice”, children received 1 point for reporting that the positively valenced puppet would be “nice”, 1 point for reporting that the negatively valenced puppet would be “not very nice”, and 1 point for reporting that the positively valenced puppet would be “nicer” (maximum score = 3). Mean scores are shown in Fig. 2 as a function of condition and trait. We first used two-tailed  $t$  tests, comparing scores for each trait with a 50% chance level of 1.5, to assess whether children selectively associated the positively valenced puppet with particular traits. In the Strength condition, mean scores on all traits exceeded chance levels, with children favoring the stronger puppet over the weaker puppet (all  $p$ s < .001). By contrast, in the Accuracy condition, mean scores exceeded chance levels for “smart” ( $t = 2.1$ ,  $p = .041$ ) but not for “strong” or “nice” ( $p$ s > .10).

To confirm these findings, we conducted a three-way repeated-measures ANOVA with condition and age group as between-participants factors and trait as a within-participants factor. In each condition, one older child was excluded due to missing one or more trait scores. This analysis revealed a significant main effect of condition,  $F(1, 54) = 15.33$ ,  $p < .001$ ,  $d = 0.84$ ; mean scores were higher among children in the Strength condition ( $M = 2.65$ ,  $SD = 0.77$ ) than in the Accuracy condition ( $M = 1.84$ ,  $SD = 1.14$ ). The effect of age approached significance,  $F(1, 54) = 3.63$ ,  $p = .062$ ,  $d = 0.39$ ; older children ( $M = 2.44$ ,  $SD = 0.96$ ) scored slightly higher than younger children ( $M = 2.04$ ,  $SD = 1.10$ ), averaging across condition and trait. The effect of trait and the Condition  $\times$  Trait interaction were not significant.

In a follow-up analysis, individual children were categorized according to whether or not they made global trait generalizations about the puppets. In the Accuracy condition, trait scores were summed for “strong” and “nice” (maximum = 6). In the Strength condition, trait scores were summed for “smart” and “nice” (maximum = 6). Using a binomial test, the chance of observing six responses favoring the positive puppet, out of six trials, is less than 5% ( $p = .016$ , one-tailed). Thus, children with scores of six were categorized as global generalizers, whereas children with scores below six were categorized as nongeneralizers. In each condition, one child was excluded due to missing a trait rating score. In total, 66% of children in the Strength condition were global generalizers ( $n = 11$  older preschoolers,  $n = 8$  younger preschoolers). In the Accuracy condition, 24% of children were global generalizers ( $n = 4$  older preschoolers,  $n = 3$  younger preschoolers). A chi-square test confirmed that a greater proportion of children in the Strength condition, as compared with the Accuracy condition, were global generalizers,  $\chi^2(1) = 10.04$ ,  $p = .002$ , Cramer's  $V = .42$ . Thus, as for the label and lifting trials and also the behavior predictions, children were prone to global generalizations about traits on the basis of physical strength but not on the basis of accuracy.

### Relationship between test trial performance and generalized evaluations

Next, to examine the overall pattern of responding, analyses were conducted to examine whether children who were selective in the label endorsement trial (Accuracy condition) or the lift endorsement trial (Strength condition) made global generalizations for both the behavior and trait ratings. Children who scored 10 or above on the behavior ratings and scored six on the trait ratings were categorized as global generalizers, whereas those with lower scores on one or both ratings were categorized as nongeneralizers. In the Accuracy condition, one child was excluded for missing both behavior and trait rating scores. In the Strength condition, of the 24 children who correctly answered the lift trial, 14 (58%) were characterized as global generalizers ( $n = 9$  older preschoolers,  $n = 5$  younger preschoolers). By contrast, in the Accuracy condition, of the 21 children who correctly answered the label trial, four (19%) were global generalizers ( $n = 3$  older preschoolers,  $n = 1$  younger preschooler). A chi-square test confirmed that a greater proportion of accurate responders in the Strength condition,

as compared with the Accuracy condition, made both behavior and trait global generalizations,  $\chi^2(1) = 7.20$ ,  $p = .007$ , Cramer's  $V = .40$ .

## Discussion

This study examined the inferences that children make when they demonstrate selective trust. We considered three inference processes that may underlie selectivity: trait-based reasoning, behavior matching, and global evaluative reasoning. Given the existing evidence against the first two processes, we begin the discussion by considering whether global evaluative reasoning guides children's selectivity. We then confirm the unlikelihood of the first two processes in light of our findings. Next, we offer another interpretation that may account for the results obtained. We end by addressing several factors that may explain the condition-related differences in the breadth of children's evaluations.

The hypothesis that global evaluations underlie selective trust predicts that children should expect an accurate informant to display positive characteristics not just in naming objects but also with respect to other behaviors and traits. The results obtained in the Accuracy condition offer no support for this prediction. As expected, children did endorse the accurate puppet rather than the inaccurate puppet when learning a novel object label, but they drew no systematic conclusions about her physical strength. With respect to behavioral domains, children in the Accuracy condition also made local inferences rather than global inferences. They predicted that the more accurate puppet would be more competent at labeling unknown objects but not at lifting objects, knowing what animals eat, throwing basketballs, or sharing cookies. Similarly, when they were explicitly asked to make trait attributions, children judged that the more accurate puppet was "smarter" but not "stronger" or "nicer".

The restricted nature of children's inferences was further underlined by three analyses of individual patterns of performance. Relatively few preschoolers in the Accuracy condition made either global behavioral predictions or global trait attributions. In addition, few children who endorsed the accurate puppet in the label task consistently generalized across behaviors and traits. These findings are consistent with those of [Brosseau-Liard and Birch \(2010, Experiment 2\)](#). They found that some (but not all) 4-year-olds identified a previously accurate labeler as likely to know more words, but they did not attribute greater factual knowledge, superior talents, or more prosocial attributes to the accurate labeler. Thus, together with the results of [Brosseau-Liard and Birch](#), the current results show that prior research on 3- and 4-year-olds' selective trust cannot be explained in terms of a global halo effect.

Should we conclude that preschoolers in the Accuracy condition engage in behavioral matching or that they attribute a trait-like disposition to accurate informants as compared with inaccurate informants? As noted in the Introduction, it is unlikely that preschoolers engage in strict behavioral matching because, in previous studies, having monitored informants for differential accuracy in naming objects, they subsequently trusted the more accurate informant in a different domain ([Birch et al., 2008](#); [Koenig & Harris, 2005](#); [Rakoczy et al., 2009](#)). In addition, several studies have shown that 3- and 4-year-olds make accurate retrospective judgments about accurate labelers as opposed to inaccurate labelers. For example, they judge that an accurate labeler was "good" at answering questions about the names of objects, whereas an inaccurate labeler was "not very good" at doing so (e.g., [Corriveau & Harris, 2009](#); [Pasquini, Corriveau, Koenig, & Harris, 2007](#)). These systematic judgments about the puppets, elicited through interview questions, require some sort of evaluation of the informants that extends beyond the immediate matching of behavior across similar contexts. Furthermore, preschoolers are capable of making such judgments up to 1 week after initial exposure to the differential accuracy of two informants ([Corriveau & Harris, 2009](#)). The persistence of children's selectivity undermines the notion that their selectivity is based on immediate matching of behaviors. In the current study, children's ratings of the accurate puppet as "smart" and the inaccurate puppet as "not smart" also suggest that they are drawing conceptual conclusions about the informants rather than simply matching similar behaviors.

Does this mean that preschoolers attribute not just a behavioral competence but also a trait-like disposition to the more accurate informant? Here the findings are mixed. In the current study, preschoolers' systematic ratings on the trait "smart" suggest that they made a relevant behavior-to-trait inference. On the other hand, they failed to predict additional factual knowledge that would be consistent with such a trait (i.e., knowing animal eating habits). The 4-year-olds interviewed by

Brosseau-Liard and Birch (2010) also failed to attribute greater factual knowledge to a more accurate puppet, whereas the 5-year-olds in the same study did make this trait-relevant generalization.

This pattern of results is consistent with studies exploring children's developing understanding of the "division of cognitive labor", that is, children's emerging ability to selectively attribute knowledge to experts (e.g., Danovitch & Keil, 2004, 2007; Keil, Stein, Webb, Billings, & Rozenblit, 2008; Lutz & Keil, 2002). For example, Lutz and Keil (2002, Experiment 1) presented 3-, 4-, and 5-year-olds with two experts: a doctor and a car mechanic. All age groups attributed relevant knowledge to the correct expert (e.g., knowing how to fix a broken arm versus a flat tire). Moreover, 4- and 5-year-olds, but not 3-year-olds, correctly attributed an understanding of the relevant underlying principles to the appropriate expert, attributing greater knowledge about biology in general to the doctor and machine functioning to the mechanic. Similarly, when children encounter an adult and a child informant, they turn to the adult over the child in domains such as novel object labeling (Jaswal & Neely, 2006), nutrition (VanderBorghet & Jaswal, 2009), and game rules (Rakoczy, Hamann, Warneken, & Tomasello, 2010), but they turn to the child over the adult in other domains such as toy functions (VanderBorghet & Jaswal, 2009). Taken together, these results suggest that children's selective attribution of expertise in some domains is already evident during the preschool years, and their awareness that expertise encompasses broader underlying principles emerges by 4–5 years of age. In our study, the puppets each represented English-speaking individuals, who are typically knowledgeable about the names of common objects. Indeed, even 16-month-olds are surprised when a person labels an object incorrectly, more so than if an incorrect label is conveyed via audio-recording (Koenig & Echols, 2003). Repeated misnaming of objects may call into question the inaccurate informant's presumed expertise about object names, leading to selective mistrust. However, like the young children in Lutz and Keil's (2002) study, the 3- and 4-year-olds in our study did not extend their distinction between the informants beyond the immediately relevant domain of word knowledge to encompass, in this case, knowing more about animals. That is, they did not infer that the accurate informant might be more knowledgeable in general. This was true both for the younger and older preschoolers in our sample. Results consistent with this proposal were also obtained by Brosseau-Liard and Birch (*in press*). In their study, 4- and 5-year-olds were familiarized with an accurate labeler and an inaccurate labeler. In subsequent test trials, the two informants made conflicting claims about the identity of an object concealed in a box. Children endorsed the claim of whichever informant had looked in the box no matter what their prior history of labeling accuracy. Thus, children did not overgeneralize prior labeling accuracy. They appropriately acknowledged the key role of visual access in establishing an object's identity. Future studies should throw more light on children's developing conceptualization of knowledgeable and less knowledgeable individuals.

So far, we have posited that repeated inaccuracies might lead a child to infer that an informant is prone to error or less knowledgeable about words. It is conceivable that children consider the inaccurate puppet to be unwilling to provide information or to be deceitful. However, the results obtained in this study are not consistent with this possibility. In particular, children in the Accuracy condition did not appear to distinguish between the puppets in terms of their prosocial or antisocial behaviors or traits; they did not identify the inaccurate informant as less likely to share cookies or as less "nice". Thus, it is unlikely that the children made a trait-like or generalized evaluation of the inaccurate puppet as deceitful or unwilling to provide known information.

In summary, when 3- and 4-year-olds monitor informants for their accuracy and subsequently prefer to learn from someone who has been more accurate, they do not do so on the basis of a halo effect (i.e., global attributions). They make more restricted assessments of their informants. Precisely how to characterize children's attributions (i.e., determining how wide-ranging the knowledge that they attribute to a more accurate informant is) warrants additional research. A plausible interpretation is that children draw on their existing assumptions about individuals as experts in object labeling in this paradigm. That is, mistakes in naming everyday objects may call into question the inaccurate informants' access to this "common knowledge". Future studies may shed further light on whether children's presumptions about object labels as common knowledge underlie their selectivity for accurate informants over inaccurate informants.

Despite the failure to find global evaluations in the Accuracy condition, children did appear to make such evaluations on the basis of demonstrations of physical strength. In the Strength condition,

children predicted that the stronger puppet would be more competent at labeling unknown objects, lifting objects, knowing what certain animals eat, throwing basketballs into a hoop, and sharing cookies. In addition, they judged the stronger puppet to be “smarter”, “stronger”, and “nicer”. The global nature of children’s inferences also emerged in the analyses of individual patterns of performance. The majority of children in the Strength condition made global behavior predictions and global trait attributions. Moreover, as compared with children who learned about relative accuracy, children who learned about relative strength were more likely to be global generalizers across both behavior predictions and trait attributions, drawing conclusions about the puppets that extended well beyond the familiarization domain of physical strength. Thus, like school-age children who make global assessments of fictional characters and peers, preschoolers make quite broad and robust attributions on the basis of demonstrations of physical strength and weakness, favoring a stronger puppet.

Thus, global evaluations appear to characterize children’s inference processes in the domain of physical strength. How can we explain the marked disparity in the breadth of generalization from accuracy information as compared with strength information? Four different explanations are considered, two focusing on the relative salience and clarity of strength and accuracy and the other two focusing on the potentially distinctive nature of inferences about strength.

In the current study, evaluations of the positive and negative informant were more polarized in the Strength condition, potentially suggesting that variation in strength was more salient to children than variation in accuracy. This type of hypothesis was suggested by Heyman, Gee, and Giles (2003), who presented preschoolers with stories about characters who performed positive behaviors in one domain but negative behaviors in another domain (nice/not-smart versus not-nice/smart). Afterward, more than half of the children misremembered the story. They were likely to misremember the nice character as being smart, but they rarely remembered the smart character as being nice. Thus, the social–moral disposition of the story character appeared to be more salient; children recalled the story in such a way that it fit their evaluation of the goodness of the character. Similarly, a halo effect might be more evident for physical strength than for accuracy because children are more likely to notice and encode variation in strength as compared with accuracy. However, existing evidence suggests that accuracy is salient to preschoolers. First, if children did not notice and encode variation in accuracy, they should not make any systematic attributions—local or more global—on the basis of accuracy. Yet, as discussed earlier, children in the Accuracy condition did make systematic local attributions. Second, preschoolers retain information about differential accuracy for up to 1 week even with no intervening reminders (Corriveau & Harris, 2009). These findings strongly suggest that differences in accuracy are salient to preschoolers and readily encoded.

A second possibility is that condition-related differences stem from variability in various features of the familiarization phases of this study. For example, the ease of inferring a specific attribute from the demonstrated behaviors may have differed; perhaps lifting is a better indicator of strength than labeling is of smartness or accuracy. As noted earlier, there are several ways to interpret inaccurate labeling, including deceitfulness and unwillingness to share good information. By contrast, the puppets’ words during the lifting attempts (i.e., “I will lift it”) temper an interpretation of the weaker puppet as unwilling to cooperate. Thus, strength, as we presented it in this study, may be a more clearly defined attribute. Competing or unclear conceptualizations of the accurate puppet may have led to a diminished halo effect in that condition. Furthermore, the puppets’ words in the Strength condition (i.e., “I will lift it”) inadvertently introduced an additional piece of information in the Strength condition; the successful lifter was also more accurate than the unsuccessful lifter in the sense that this claim was true for her but false for the other puppet. It is not clear whether this feature of the familiarization phase amplified the distinction between the stronger and weaker puppets. Although the familiarization phase for both attributes demonstrated more positive and more negative behavior, future studies should consider the content of the familiarization phase more closely.

A third possible explanation of the results is that children in the Strength condition, but not in the Accuracy condition, made an inference that was linked to the assumed age of the informant. On the basis of the lifting familiarization trials, children may have inferred that the successful lifter was older than the unsuccessful lifter. Using this characterization, they may have gone on to prefer the “older” puppet’s label and to assume that she would be more successful at the various behavioral tasks. However, recall that children also claimed that the stronger puppet would share more and was nicer. There

is no obvious reason why children would make such attributions to someone perceived to be older. In any case, the puppets were deliberately chosen to be similar in appearance so as to reduce the potential for any bias related to superficial differences. Nevertheless, in future research, information about the age of the two target puppets could be provided explicitly to rule out the possibility that children use perceived age as a basis for inference.

Turning to the fourth possible interpretation, it is plausible that the disparity in the breadth of children's inferences arises from differences in accuracy and strength as types of competencies. Labeling accuracy implies a specific knowledge-based capacity, as does the expertise of a doctor or mechanic. By contrast, physical strength conveyed through lifting success may be a more general behavioral disposition. It does not rely on access to any particular form of informational expertise. Thus, the reasoning processes involved in assessing accuracy may be qualitatively different from those involved in assessing strength. In particular, children's early emerging ability to reason about individuals as experts in varied domains may facilitate their selectivity in the case of object labeling accuracy but not in the case of object lifting success. Outside the realm of specific knowledge-based capacities, children may default into making global evaluations.

In conclusion, research using the selective trust paradigm has provided robust evidence that children distinguish between competing informants on the basis of their prior accuracy in labeling known objects. One hitherto plausible interpretation of that selectivity is that preschoolers are prone to a halo effect; they regard a display of relative competence as an index of globally positive characteristics. The current findings show that 3- and 4-year-olds are indeed prone to such a halo effect when familiarized with individuals distinguished by their physical strength. However, we found no evidence of such global generalization when young children encounter accurate versus inaccurate informants.

## Acknowledgments

We thank Vanita Srikanth and Tara Chiatovich for their assistance with data collection as well as the staff, volunteers, and visitors of the Boston Museum of Science Discovery Center.

## References

- Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? An analysis of children's behavioral predictions. *Child Development, 72*, 1409–1425.
- Astington, J. W. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind* (pp. 157–172). Hillsdale, NJ: Lawrence Erlbaum.
- Benenson, J., & Dweck, C. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development, 57*, 1179–1187.
- Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018–1034.
- Brosseau-Liard, P. E., & Birch, S. A. J. (2010). "I bet you know more and are nicer too!" What children infer from others' accuracy. *Developmental Science, 13*, 772–778.
- Brosseau-Liard, P. E., & Birch, S. A. J. (in press). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development*.
- Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers' ability to make dispositional predictions within and across domains. *Social Development, 6*, 53–75.
- 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.
- 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.
- Danovitch, J. H., & Keil, F. C. (2007). Choosing between hearts and minds: Children's understanding of moral advisors. *Cognitive Development, 22*, 110–123.
- Heyman, G. D., Gee, D. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development, 74*, 516–534.
- Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology, 77*, 1–19.
- 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.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.
- Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition, 87*, 179–208.

- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277.
- Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development, 78*, 1543–1558.
- Livesley, W. J., & Bromley, D. B. (1973). *Person perception in childhood and adolescence*. New York: John Wiley.
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216–1226.
- Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73*, 1073–1084.
- Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology, 28*, 785–798.
- Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development, 24*, 61–69.
- Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*, 550–560.
- Rotenberg, K. J. (1980). Children's use of intentionality in judgments of character and dispositions. *Child Development, 51*, 282–284.
- Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*, 669–679.
- Stipek, D. J., & Daniels, D. H. (1990). Children's use of dispositional attributions in predicting the performance and behavior of classmates. *Journal of Applied Developmental Psychology, 11*, 13–28.
- VanderBorgh, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development, 18*, 61–71.
- Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harns (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis: University of Minnesota Press.
- Yuill, N. (1993). Children's understanding of traits. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 87–110). New York: Guilford.