Developmental Psychology
2015, Vol. 51, No. 11, 1574–1587

0012-1649/15/$12.00 http://dx.doi.org/10.1037/dev0000042

The Role of Trait Reasoning in Young Children's Selective Trust

Jonas Hermes, Tanya Behne, and Hannes Rakoczy
University of Göttingen

In recent years, ample research has shown that preschoolers choose selectively who to learn from, preferring, for example, to learn novel words from a previously accurate over a previously inaccurate model. But this research has not yet resolved what cognitive foundations such selectivity builds upon. The present article reports 2 studies that investigate whether preschoolers' selective trust is based on global impression formation (halo effects), on the close matching of past and future behavior or on trait-based inferences, and moreover whether the cognitive strategies used are the same for both positive and negative information (high and low competent models). Four- and 5-year-old children ($N = 96$) were presented with 2 high-competence models (strong vs. knowledgeable; Study 1) or 2 low-competence ones (weak vs. unknowledgeable; Study 2). In 5 subsequent task groups, which required strength and knowledge to different degrees, children were asked to choose between the 2 models. Children in both studies chose models selectively in accordance with their corresponding attributes, preferring the strong (or avoiding the weak) model for strength-related tasks and preferring the knowledgeable (or avoiding the unknowledgeable) model for knowledge-related tasks. This pattern of selective model choice held only for those children who correctly identified the attributes of both models (strong, smart), as indicated by their answers to trait questions at the end of the session. This suggests that trait reasoning plays a crucial role in young children's selective social learning.

Keywords: selective trust, social learning, trait ascription, social cognition, inductive inferences

When growing up, children need to acquire a wealth of information, much of which can only be gained socially from others. In the past decade, much developmental research has investigated whether children acquiring such information from others simply trust everyone alike or whether and how they are selective in whom they trust (for reviews, see Harris, 2007; Mills, 2013).

In research on selective trust,[1] the child is generally presented with two models that vary in a certain respect. When then confronted with a novel task or problem, the child can seek help, accept information from or attribute competence to one of the two models. For example, Koenig and Harris (2005, Exp. 1) presented two models that varied in the accuracy with which they labeled familiar objects: One constantly labeled them correctly whereas the other always labeled them incorrectly (e.g., calling a ball "cup"). In the test phase, unknown objects were presented for which both models provided novel labels. Children from age 4 selectively endorsed the labels from the previously competent model. Subsequent research has shown that children are selective with respect to numerous model attributes, preferring to learn, for example, from models that are nicer (Landrum, Mills, & Johnston, 2013), more similar or familiar to them (Corriveau & Harris, 2009; Kinzler et al., 2011), or that expresses certainty rather than uncertainty (Sabbagh & Baldwin, 2001).

Cognitive Foundations of Selective Trust

What are the cognitive foundations of such selective social learning? Theoretically there are three possibilities (see Figure 1 and Fusaro, Corriveau, & Harris, 2011). First, children might engage in behavior-matching, drawing narrow inferences on a behavioral basis only, closely matching past and future behavior without attributing any competence to the model. For example, someone who had been good at labeling objects in the past would be expected to be good at labeling objects in the future but would not be expected to be competent with respect to related behaviors. Empirically, however, numerous studies have indicated that even young children draw much wider generalizations from the behavior they have witnessed, including generalizations that cannot be explained by behavior matching (e.g., from labeling to knowledge about functions; Koenig & Harris, 2005, Exp. 3). In fact, research

This article was published Online First September 21, 2015.

Jonas Hermes, Tanya Behne, and Hannes Rakoczy, Department of Developmental Psychology, Institute of Psychology, University of Göttingen.

Thank you very much to Marlen Kaufmann and Konstanze Schirmer for help with recruiting children and organizing the study. We also thank all the children who participated as well as their parents and preschool teachers. This work was supported by the German Science Foundation, Grant Ra-2155/1–1.

Correspondence concerning this article should be addressed to Jonas Hermes, Department of Developmental Psychology, Institute of Psychology, University of Göttingen, Waldweg 26, D- 37073 Göttingen, Germany. E-mail: jonas.hermes@psych.uni-goettingen.de

[1] We follow recent developmental research (e.g., Harris, 2007, 2012; Harris & Corriveau, 2011; Kinzler, Corriveau, & Harris, 2011; Over, Carpenter, Spears, & Gattis, 2013; Robinson & Einav, 2014) in using "selective trust" in a very broad sense as an umbrella term covering all kinds of selective recruitment of one of several models for informational purposes (who to learn from, who to believe, who to acquire information from, etc.) as well as for practical purposes (who to use as a cooperative partner, etc.).

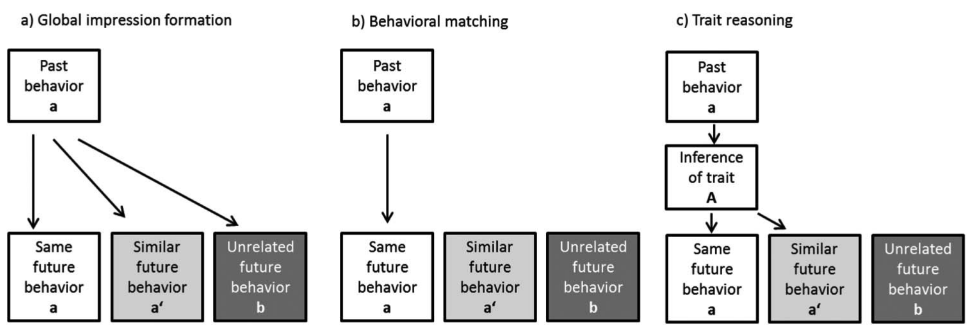


Figure 1. Scopes of generalizations under three possible patterns of cognitive foundations that may underlie preschoolers' selective trust.

has shown that preschoolers sometimes generalize overly widely, to completely unrelated domains, expecting, for example, a knowledgeable model to be nice (Brosseau-Liard & Birch, 2010, Exp. 1), a nice model to be smart and athletic (Cain, Heyman, & Walker, 1997), a strong model to be nice and smart, (Fusaro et al., 2011) or an attractive model to be knowledgeable (Bascandziev & Harris, 2014).

These findings seem more compatible with a second possibility: that children's selective learning is based on global impression formation. According to this possibility, children perceive someone introduced as competent in a certain domain (e.g., naming objects) as positive in some global, undifferentiated way (in the style of halo effects) and would thus selectively recruit her for both similar and totally unrelated problems. Similarly, children would consider someone introduced as incompetent in a certain domain as globally negative and avoid her in similar but also in totally unrelated domains (pitchfork effect).

The third possibility, finally, is that children engage in more sophisticated reasoning: based on the observation of a model's (in-)competent behavior in a certain domain and on their conceptual background knowledge, children ascribe certain attributes to this model (e.g., being smart, knowledgeable about XYZ or strong) upon which they base their inferences regarding future behavior of the model and their decision whom to recruit for a given task with specific requirements ("this task requires knowledge about XYZ, so I should ask . . ."; Sobel & Kushnir, 2013). Empirically, a number of studies show patterns of selective social learning in line with this suggestion. For example, in a study by VanderBorght and Jaswal (2009), children showed sensitivity concerning the domain of competence in their selective trust. They trusted adults over children in a domain where adults usually know more (food questions) but trusted children over adults in a domain where children usually know more (toy questions). Lutz and Keil (2002) showed that even 3- and 4-year-old children understand the division of cognitive labor. When they encountered experts from different domains (e.g., a doctor and a car mechanic) and were confronted with questions from these different domains of expertise, children reasonably attributed knowledge to the respective expert. Similarly, Kushnir, Vredenburgh, and Schneider (2013) introduced two different experts: one competent in fixing toys and one competent in labeling toys. In subsequent test trials, children selectively requested help from the toy fixer when needing to fix a toy, and from the competent labeler when trying to find out the name of a toy. These studies thus suggest that even preschool children ascribe certain attributes to a model, based on his or her past behavior, and recruit this model selectively when confronted with a problem that lies within the scope of the model's competence.

Selective Trust and Trait Reasoning

Such an interpretation might seem surprising, however, in light of the findings of a different line of research: children's developing trait ascription. This research, though using very similar methods, has generally found that trait ascription emerges only at a much later age. In a common paradigm in this field, children see protagonists who show a certain behavior (e.g., helping) which is an expression of a certain trait (e.g., prosociality). In a following test phase, children are then asked to predict the protagonists' future behavior. In these studies, children usually do not engage in competent, adult-like inferences from observed behavior to similar future behavior until their late primary school years (e.g., Rholes & Ruble, 1984; see Heyman, 2009; Rholes, Newman, & Ruble, 1990 for reviews). Rather, younger children have been found to overgeneralize trait-relevant information in broad and undifferentiated ways—which has been interpreted as evidence that children do not attribute traits in adult like ways before age 9 or 10.

So how do the results in research on selective trust and trait ascription relate to one another? One possibility is that the findings of the trait ascription literature cast doubt on the claim that preschoolers use trait reasoning in their selective trust. Alternatively, however, the later competences found in the trait ascription literature might be a result of the specific task demands and performance factors of the kinds of tests used in this research that might mask actual competences in younger children. Along such lines, Liu, Gelman, and Wellman (2007) argued that in trait ascription tasks, the whole inference process required to predict future behavior on the basis of past behavior can be divided in two parts: (a) inferring a trait from past behavior and (b) predicting future behavior on the basis of a trait. And in fact, Liu et al. (2007) have shown that although children from age 4 were not able to combine both processes (i.e., to draw behavior-to-behavior inferences), they competently engaged in each single part of the process (behavior-to-trait or trait-to-behavior inferences). Similarly, Li, Heyman, Xu, and Lee (2014) provided preschoolers with explicit trait information and found that by 5 years of age children differentiated rationally between trust-relevant and trust-irrelevant measures in their trust judgments. Furthermore, studies using alternative, nonexplicit measures like the anticipation of emotions or the inference of mental states provide additional evidence for earlier forms of trait understanding (see Heyman, 2009). For example, Heyman and Gelman (1999) showed that 4-year-olds can use trait information to make inferences about mental states, expecting, for example, that a shy person would be less pleased when encountering many people compared to a person who is not shy.

Along similar lines, Sobel and Kushnir (2013) have recently proposed a theoretical account of children's developing selective social learning that aims at reconciling seemingly contradictory findings. The basis claim of this account is that children engage in rational selective social learning if and to the degree that they have available the requisite conceptual knowledge about the target domain and the model attributes in question. Only when no such knowledge is available, the account claims, will children revert to more unspecific general strategies, such as using valence in halo-like ways. The prediction—so far not tested empirically—follows that whether children engage in rational trait-based selective learning will depend on their conceptual grasp of the traits in question: When children have available a given trait concept in principle, and when they can apply it in the specific case at hand, they will engage in rational selective learning; otherwise they will use simpler strategies.

Positive and Negative Information

All in all, thus, much recent research has shown that young children engage in selective social learning but it leaves open what the cognitive foundations and underlying processes are. A related open question is whether the same cognitive foundations and processes are in play when dealing with positive information (competence) and negative information (incompetence). Although most studies have contrasted same versus lower competence models (relative to the child's competence) and sometimes same versus higher competence models, few studies have systematically contrasted higher and lower competence models. And the few studies that have done so, interestingly, suggest that in fact different kinds of processes might underlie children's dealing with competence versus incompetence: First of all, preschool children might understand above average competences slightly later than incompetence (Corriveau, Meints, & Harris, 2009). Furthermore, one recent study suggests that children might apply processes of global impression formation more when dealing with information about incompetence compared to competence (Koenig & Jaswal, 2011): Children preferred a dog expert (precisely knowing dog breeds, e.g., "That's a Basenji Dog") over a nonexpert (knowing only visible features, e.g., "That's a black dog") for information about dogs, but not artifacts. However, when an inexpert (wrong claims, e.g., "That's a Basenji Cat") was contrasted with a neutral informant (only drawing attention to objects: "That's a nice one"), children preferred the neutral informant over the inexpert for both dogs and artifacts. What this study suggests, thus, is that although there may be no halo-effect for perceived competence in children's selective learning, there may well be a pitchfork effect for perceived incompetence.

The Present Studies

Against this background of mixed findings and open questions, the rationale of the present studies was to systematically investigate the cognitive foundations of early selective trust and model choice: First of all, are early selective learning and trust based on rational trait reasoning rather than on simpler processes such as global impression formation or narrow behavior matching? And are the same kinds of processes in play in dealing with positive (competence) versus negative (incompetence) information? Methodologically, we pursued these questions in the following way: First, we presented children with contrast pairs of models most suitable for uncovering the cognitive foundations of selective model choice, namely models with comparable competence, yet in different domains (e.g., physically strong/verbally accurate). A number of previous studies have tried to investigate the cognitive processes underlying early selective trust with different contrast cases: Children were presented with pairs of models that varied in competence on one dimension (e.g., reliable vs. unreliable) and were then given the choice between these models in tasks for which the competence in question was relevant or irrelevant (Brosseau-Liard & Birch, 2010; Fusaro et al., 2011). The results from these studies show that in some cases children generalized widely in halo-like ways from a given competence to relevant and irrelevant domains and thus seem to speak against rational trait ascription and in favor of global impression formation. For example, in a study with a very similar approach as the present one, Fusaro and colleagues (2011) presented children with a pair of agents that varied in physical strength (strong-weak) or a pair of agents that varied in verbal accuracy (accurate-inaccurate). In subsequent tasks, children were asked to endorse one of two conflicting labels for a novel object provided by each agent or to judge who of the two agents successfully lifted an object. The results showed that children widely generalized from physical strength in halo-like ways, preferring to learn novel words from a strong rather than from a weak model. A fundamental problem with the kinds of contrasts used in these studies (between agents differing in competence in one and the same dimension), however, is that they might produce false negatives and mask children's actual competence for rational trait ascription. Children might well be capable in principle of rational trait-based inferences, yet fall back on simpler strategies such as global impression formation or related heuristics (e.g., "just take the one that was good") as long as they yield determinate answers. To rule out such strategies, in the present study we thus used contrast pairs of models that were equally good, yet in different domains. Second, we tested for children's inductive generalizations not only in the settings in which the model competences (strength/accuracy) were explicitly introduced, but in more systematic and fine-grained ways in a number of different types of tasks that required competence in these domains to different degrees. Third, to test specifically for the relation between trait ascription and selective learning, we assessed children's knowledge of the models' traits in question explicitly ("who is stronger / smarter?") and correlated it with children's selective model recruitment. Fourth, we presented children with contrast pairs of two models that were either equally good in different dimensions (Study 1) or equally bad (Study 2).

The design of the present studies thus implements a stringent test between the theoretical possibilities concerning children's selective trust given that they predict clearly distinct response patterns (see Figure 2): If global impression formation were the basis of preschoolers' selective trust, we would expect no selectivity in any of these tasks because both models showed high (or low) competence and in case of global impression formation only the degree but not the domain of competence would be crucial. If children closely matched past and future behavior, selective model choice would occur only in tasks in which the requirements matched the demonstrated behavior. Finally, if children engaged in trait reasoning, they should also be selective in tasks that predominantly require strength or knowledge. And in line with the prediction of the rational selective learning account by Sobel and Kushnir (2013), there should be a clear relation between their performance on trait ascription tasks and their performance on selective model choice tasks: those children competent at ascribing the relevant traits (e.g., "strong," "smart") should perform systematically and accordingly in their selective model choice.

Study 1 (High-Competence Models)

In Study 1 we presented preschoolers with a pair of highly competent models: Model 1 was shown to have above-average competence in labeling objects and Model 2 to have above-average competence in lifting objects. In the test phase, children were then asked to choose between these models in tasks that require one of the two competencies to varying degrees. When forced to choose between the two models in a task requiring mainly smartness, for example, the child has positive information about the high competence in question concerning Model 1, but no relevant information concerning Model 2. In this context, a rational strategy would be to assume that in the absence of relevant information, Model 2 should likely correspond to the default, that is, to have average competence on the dimension in question (smartness) and thus to prefer Model 1. We predicted that children would apply this rational strategy and would choose Model 1 for all tasks predominantly requiring smartness and Model 2 for all tasks predominately requiring strengths.

Method

Participants. Children from mixed socioeconomic backgrounds were recruited in a medium-sized German town from a database of families who had agreed to participate in developmental studies. Twenty-four 4-year-old children (age range: 48–59 months, M = 53 months, SD = 3.5 months, 14 girls) and 24 5-year-old children (age range: 62–71 months, M = 66 months, SD = 2.9 months, nine girls) were included in the final sample. Eleven additional children were tested but excluded from analysis because of failure to cooperate (n = 3) or failure to answer the comprehension questions (n = 8). Children were tested individually either in their day care center or in the child lab. Test sessions lasted approximately 20 min.

Material and procedure. The test session started with two demonstration phases (strength demonstration and accuracy demonstration, order counterbalanced), followed by two test blocks (order counterbalanced), with a demonstration reminder in between the test blocks, and it ended with four trait and preference questions. (All material was shown as a computer presentation that combined still scenes and embedded video clips.)

Demonstration phases. Children encountered two pairs of puppet models, one pair that differed in strength and one that differed in accuracy. The four puppets were each dressed in a different color and were named accordingly (e.g., "Ms. Green").

Accuracy demonstration phase. First, children were introduced to two puppet models. Then in each of the four subsequent trials, these two models were presented with a picture of a known object (e.g., an airplane, see Appendix A). The experimenter commented "oh, look what they have" and showed the child a high resolution print of the same object, asking "Do you know what this is?" Children generally provided the common label (e.g., airplane). The experimenter double-checked on his list and provided an expert label for the object (e.g., supersonic airplane). The children were then shown a video in which both puppets provided different labels for the object (order counterbalanced). One puppet constantly provided the same expert labels previously read from the list whereas the other puppet constantly provided wrong labels (e.g., "helicopter"). Expert labels always included the label commonly used for the object (e.g., "supersonic airplane" for "airplane"), so that they would not be perceived as inaccurate. Wrong labels were part of the same higher order category (e.g., "helicopter" for "airplane") because the inacurate model was supposed to be perceived as inaccurate but not as bizarre or deceptive. The aim was to present the expert model's knowledge as more elaborate than the child's knowledge and to present the inaccurate model's knowledge as less elaborate than the child's. After each trial the experimenter repeated what both puppets said. At the end of this phase, the experimenter asked two comprehension questions: "Who was good/Who was not so good at labeling these things, Ms. . . . or Ms. . . .?" If the child did not correctly answer both

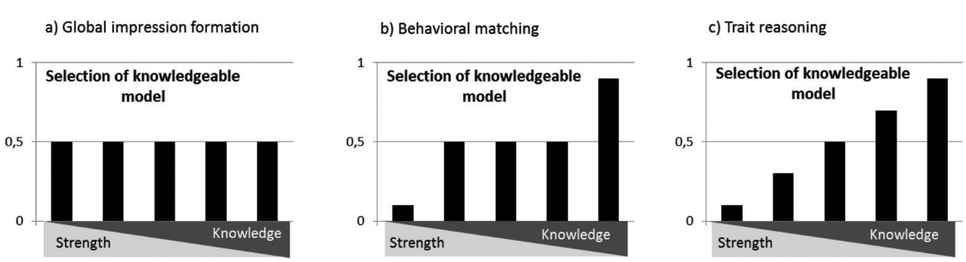


Figure 2. Expected results based on different cognitive foundations of selective trust. (Required model behavior in high-strength and high-knowledge tasks closely resembles behavior shown in the demonstration phase.)

questions, an additional demonstration trial was presented and the same questions were repeated. Children who still did not correctly identify the puppets were excluded from analysis (see Appendix D). After the comprehension questions of the first demonstration block, the experimenter explained that the two puppets had to go now and immediately introduced the two other puppets.

Strength demonstration phase. After both puppets were introduced, they were each presented with an object lying in front of them. One puppet had a heavy object (e.g., a big suitcase), whereas the other puppet had a light object (e.g., a small story book; see Appendix B for details). The experimenter started the video, in which the puppet with the heavy object succeeded in lifting it, saying "I'm good at that" whereas the other puppet tried but failed to lift the light object, saying "I can't manage this" (order counterbalanced). The experimenter repeated the events saying "Ms. . . . was able to lift the big suitcase but Ms. . . . wasn't able to lift the small story book." With this procedure, we intended for the child to perceive one puppet as stronger and the other as less strong than herself. After four demonstration trials, we asked two comprehension questions: "Who was good/Who was not so good at lifting these things, Ms. . . . or Ms. . . .?" All children in the final sample answered the questions correctly, so that no additional demonstration trials were needed.

Model rotation. After the second demonstration block, the incompetent model left and the competent model from the first demonstration block reappeared. For the rest of the test session, the strong and the accurate puppet were presented together. Two additional comprehension questions were posed one referring to each model (e.g., "Ms. Green, was she good or not so got at lifting things?" and "Ms. Red, was she good or not so got at labeling things?"). If the child did not answer correctly, two additional demonstration trials were presented (the accurate puppet labeling one more object and the strong puppet lifting one more object) and the comprehension questions were repeated. If a child still did not remember correctly, she was excluded from analysis (see Appendix D).

Test blocks. Two test blocks were presented (order counterbalanced). In each test trial, the models were presented with an object and the experimenter said, "Look what they have! I brought the same for you" and gave the object to the child to act on and explained it, if necessary. After the child was familiar with the object, a test question was asked in which she was invited to choose the more competent model (see Appendix C for objects, explanations, and exact test questions).

Test Block A consisted of two task blocks: one block with knowledge tasks and one block with strength tasks (three trials each, order of task blocks counterbalanced). For the knowledge tasks, the models were presented with a novel, unknown object and children were asked whether they knew the object. If they guessed incorrectly, the guesses were doubted by the experimenter. If a child actually knew the object, another object was introduced. Subsequently both puppets provided different artifact labels for the novel object (e.g., "That's a Mido, yes that's a Mido"/"That's a Toma . . ."). The experimenter repeated what the puppets said and as a crucial test question asked the child what she thought the object was. If the child did not choose between the labels provided by the puppets, the experimenter repeated the labels again and added " . . . who is right?" For the strength tasks, the models were presented with objects that required physical strength to act on (e.g., a heavy brick that needed to be carried a long distance). The child was given the object to act on and was asked which of the models would be able to perform a certain action with them. Rarely, children first chose both models. When this did occur, children were asked again to decide ("Who would be better than the other?").

Test Block B consisted of three classes of problem solving tasks (two trials each, order of task groups counterbalanced). For the problem-solving strength tasks, children needed strength, but also some skillfulness to act on these objects (e.g., opening a tight knot). For the problem-solving neutral tasks, children needed some dexterity and cognitive ability, but the tasks were neutral with respect to strength and object knowledge (e.g., putting together a puzzle cube). The problem-solving knowledge tasks followed the same procedure as the knowledge tasks (see above), but the puppets provided no labels and the children were asked which of the puppets would know what the object is good for. These tasks required some object knowledge but no strength.

Demonstration reminder. A demonstration reminder was presented between both test blocks. In two video clips the strong puppet lifted a heavy object, in the other two clips the accurate puppet provided an expert label for a known object (the procedure was the same as described in the initial demonstration phases; for details, see Appendices A and B). Finally the same comprehension questions as after the puppet rotation were asked. If a child answered incorrectly, two more demonstrations (one of each model) were shown and the questions were repeated. If a child still answered these incorrectly, she was excluded from analysis (see Appendix D).

Trait questions and preference questions. At the end, we asked two trait questions ("Who is stronger, [Ms. Green] or [Ms. Blue]?" and "Who is smarter . . .?) and two preference questions ("Who is nicer . . .?" and "Who would share her sweets with other children . . .?"). One child in the final sample ended the session before answering these questions.

Coding procedure. One coder watched and coded all video recordings. We accepted pointing or saying the puppets' name or color as answers. A "1" was coded when the accurate puppet was chosen and "0" was coded when the strong puppet was chosen. If children answered "none" or "both", we coded chance level (0.5). In the test trials, two single data points were missing because children provided no answer. We coded chance level (0.5) for these trials. As dependent variables for subsequent statistical analysis, we calculated the mean proportion of children's model choice by dividing the sum score by the number of trials for each of the five task groups.

Results[2] and Discussion

Model choice. We were interested in whether and to what degree children chose the models selectively in accordance with the individual model competences and the requirements of the tasks. A repeated-measure analysis of variance (ANOVA) on the children's model choices with age group as between-subjects-factor and task as within-subject-factor yielded a main effect of task, $F(4, 184) = 10.75$, $p < .01$, partial $\eta^2 = .19$, and neither an age effect nor an interaction. Children's model choices thus varied as a function of task. As shown in Figure 3a, children preferred the accurate over the strong model in the knowledge tasks, $t(47) =$

[2] The results of the parametric analyses reported in the following were confirmed by the results from nonparametric tests (with a single exception indicated below).

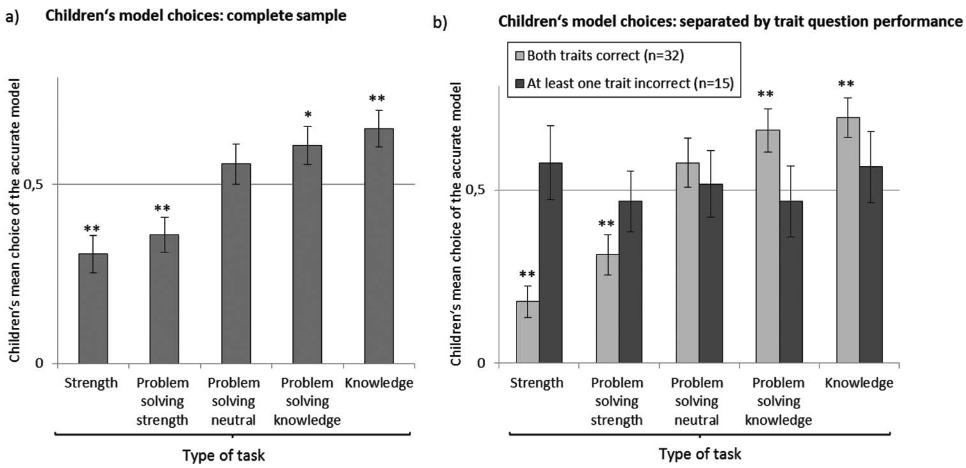


Figure 3. Children's mean choices for the accurate (instead of the strong) model for the five types of tasks for the whole sample (3a) and separated by trait-question performance (3b). Asterisks show deviation from chance level (one-sample *t*-tests, * $p < .05$, ** $p < .01$).

3.08, $p < .01$, $d = .46$ and the problem-solving knowledge tasks,[3] $t(47) = 2.04$, $p < .05$, $d = .30$, whereas they preferred the strong over the accurate model in the strength tasks, $t(47) = 3.77$, $p < .01$, $d = .54$ and the problem-solving strength tasks, $t(47) = 2.92$, $p < .01$, $d = .42$. However, children chose models randomly in the problem-solving neutral tasks, $t(47) = 1.03$, $p = .31$, $d = .15$.

Preference and trait questions. Children's model choices in the preference and trait questions were compared to chance using one sample Wilcoxon signed-ranks test against the median 0.5 (code for indifference answer: both/none). Children's choices did not differ from chance when asked who was nicer (accurate model: 51%, strong model: 32%, both/none: 17%; $Z = 1.44$, $p = .15$, $r = .21$) or who would share her sweets (accurate model: 53%, strong model: 43%, none/both: 4%; $Z = 0.75$, $p = .46$, $r = .11$). In contrast, regarding the trait questions, children's pattern of answers significantly differed from chance: They chose the previously strong model when asked who was stronger (accurate model: 21%; strong model: 77%; both/none: 2%; $Z = 3.83$, $p < .01$, $r = .56$) and the previously accurate model when asked who was smarter (accurate model: 75%, strong model: 23%, both/none: 2%; $Z = 3.54$, $p < .01$, $r = .52$).

Model choice by trait ascription. Further we analyzed whether children's selectivity in model choices depended on correct trait ascription. To this end, we created a new variable—trait question performance—that was coded "1" if both traits were ascribed correctly ($n = 32$) and "0" if at least one mistake was made in trait ascriptions ($n = 15$). Figure 3b depicts children's model choices for each task as a function of trait question performance. A repeated-measure ANOVA on the children's model choices with trait question performance and age group as between-subjects factors and task as within-subject factor revealed a main effect of task, $F(4, 172) = 5.36$, $p < .01$, partial $\eta^2 = .11$, and an interaction between task and trait question performance, $F(4, 172) = 6.22$, $p < .01$, partial $\eta^2 = .13$. Next, we compared performance in the different tasks against chance separately as a function of trait question performance. The results show that the pattern of selective model choices shown in the whole sample (Figure 3a) only holds for those children who answered both trait questions correctly—$ts(31) > 2.77$, $ps < .01$, $ds > .48$ for strength tasks, problem solving strength tasks, knowledge tasks and problem solving knowledge tasks; $t(31) = 1.09$, $p = .28$, $d = .20$ for problem solving neutral tasks—whereas children who made at least one mistake in their trait ascriptions selected models at chance level in all task groups, $ts(14) < 0.73$, $ps > .47$, $ds < .20$.

In sum, the results from Study 1 show that children recruited models in accordance with the models' competences, preferring the strong model for strength-related tasks and the accurate model for knowledge related tasks. Further analysis of the data split by trait-question performance showed that selective model choice depended on correct trait ascription, as the reported pattern only holds for children who identified the traits correctly whereas children who did not correctly identify model traits did not choose models selectively. This pattern of results supports the idea that children's selective trust is based on trait reasoning.

One open question, though, is how robust and general such trait-based rational selective trust is, in particular concerning behavior with negative rather than positive valence. Recent evidence suggests that children, suffering from a kind of negativity bias, might appear to be more rational when confronted with a model revealing positive as compared to negative behavior. When witnessing two models who, say, differed in their labeling accuracy—one on the child's level of accuracy, the other one an expert—children appropriately preferred the expert model for tasks revolving around knowledge in the according domain but did not reveal any halo effects (i.e., did not overgeneralize and prefer this model in unrelated domains). In contrast, when witnessing a model neutral with respect to labeling accuracy, and another one labeling things less accurately than the child herself, children revealed a "pitchfork effect," avoiding to learn from the less accurate labeler in all kinds of related and unrelated domains (Koenig & Jaswal,

[3] This difference failed to reach significance when we calculated a nonparametric Wilcoxon signed-ranks test ($Z = 1.9$, $p = .06$).

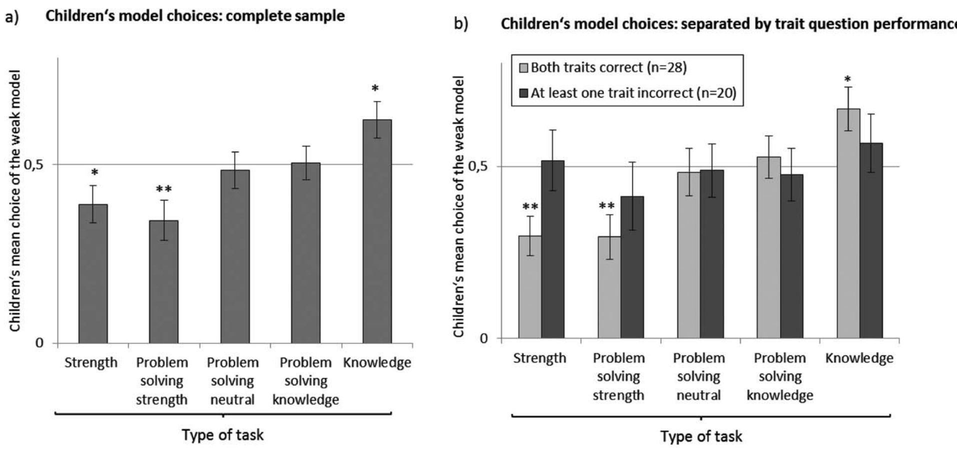


Figure 4. Children's mean choices for the weak (instead of the inaccurate) model for the five types of tasks for the complete sample (a) and separated by trait-question performance (b). Asterisks show deviation from chance level (one-sample *t* tests, * $p < .05$, ** $p < .01$).

2011). Against this background, Study 2 tested for trait-based rational selective trust with the same material and general procedure as in Study 1, but confronting the children with two low competent models (one weak, one inaccurate) in the test phase.

Study 2 (Low-Competence Models)

Method

Participants. Children were recruited from the same database as in Study 1. Twenty-four 4-year-old children (age range: 48–59 months, $M = 53$ months, $SD = 3.5$ months, 10 girls) and 24 5-year-old children (age range: 62–71 months, $M = 67$ months, $SD = 2.1$ month, 12 girls) were included in the final sample. Nine additional children were tested but excluded from analysis because of failure to cooperate ($n = 1$) or failure to answer the comprehension questions ($n = 8$).

Material and procedure. The materials and procedure were exactly the same as in Study 1 with only two exceptions: First, for the model rotation after the last trial of the second demonstration phase, the competent model left and the incompetent model from the first demonstration phase returned. The two remaining incompetent models, the inaccurate model and the weak model, remained for the rest of the test session. Second, as a demonstration reminder, children saw two video clips with the weak puppet failing to lift a light object and two video clips with the inaccurate puppet providing an incorrect label for a known object (for the lists of objects used, see Appendixes A and B). For more information on the number of children who provided incorrect answers to comprehension questions and hence needed extended demonstrations or were excluded from analysis see Appendix D. The coding procedure followed the same rationale as in Study 1. A "1" was coded when children chose the weak puppet and "0" was coded when children chose the inaccurate puppet. If children answered "none" or "both," this received the code 0.5. As dependent variables for subsequent statistical analysis, we calculated the mean proportion of children's model choice by dividing the sum score by the number of trials for each of the five task groups.

Results[4] and Discussion

Model choices. We were interested whether children selected models selectively in accordance with the individual model competences and the requirements of the tasks even when models show low competences (weak and inaccurate). A repeated-measure ANOVA on children's model choices with age group as between-subjects factor and task as within-subject factor revealed a main effect of task, $F(4, 184) = 6.13$, $p < .01$, partial $\eta^2 = .12$, but neither an age effect nor an interaction. This shows that children's model choices again varied as a function of task. As depicted in Figure 4a, children preferred the weak over the inaccurate model in the knowledge tasks, $t(47) = 2.44$, $p < .05$, $d = .37$, but preferred the inaccurate over the weak model in the strength tasks, $t(47) = 2.14$, $p < .05$, $d = .31$, and the problem solving strength tasks, $t(47) = 2.81$, $p < .01$, $d = .42$. Children chose models randomly in the problem-solving neutral tasks, $t(47) = 0.31$, $p = .72$, $d = .06$, and the problem-solving knowledge tasks, $t(47) = 0.11$, $p = .91$, $d = .03$.

Preference and trait questions. Children's model choices in the preference and trait questions were compared to chance by using one sample Wilcoxon signed-ranks test against the median 0.5, the code for indifferent answer (both/none). Children's choices did not differ from chance when asked who was nicer or who would share her sweets (for both questions: weak model: 46%, inaccurate model: 42%, none/both: 13%; $Z = 0.31$, $p = .76$, $r = .04$). In contrast, on both trait questions, children's choices differed significantly from chance: they, chose the previously inaccurate (thereby avoiding the weak) model when asked who was stronger (weak model: 17%, inaccurate model: 81%, both/none: 2%; $Z = 4.52$, $p < .01$, $r = .65$) and they chose the previously weak (thereby avoiding the inaccurate) model when asked who was smarter (weak model: 75%, inaccurate model: 23%, both/none: 2%; $Z = 3.65$, $p < .01$, $r = .53$).

[4] Again, these results were replicated with nonparametric statistics.

Model choice by trait ascription. To further analyze whether children's selective model choices depended on correct trait ascription, we created the same variable—trait question performance—as in Study 1. Figure 4b shows children's model choices in the different tasks separately as a function of this variable. A repeated-measure ANOVA on children's model choices with trait question performance and age group as between-subjects factors and task as within-subject factor was conducted. We found a main effect of task, $F(4, 176) = 4.97$, $p < .01$, partial $\eta^2 = .10$, and a trend for an interaction between task and trait question performance, $F(4, 176) = 1.96$, $p = .10$, partial $\eta^2 = .04$. Subsequent tests of children's performance in the different tasks against chance, conducted separately as a function of trait question performance, revealed that the pattern of selective model choices shown in the complete sample only holds for children who answered both trait questions correctly: $ts(27) > 2.59$, $ps < .05$, $ds > .49$ for strength tasks, problem solving strength tasks and knowledge tasks; $ts(27) < 0.45$, $ps > .66$, $ds < .10$ for problem solving neutral tasks and problem solving knowledge tasks. In contrast, children who made at least one mistake in their trait ascriptions selected models at chance level in all task groups, $ts(19) < 0.90$, $ps > .38$, $ds < .20$.

In sum, the results from Study 2 closely resemble those from Study 1. Children avoided the weak model in the strength and the problem-solving strength tasks and they avoided the inaccurate model in the knowledge tasks (but not in the problem-solving knowledge tasks). Thus, they did not engage in global impression formation but showed a similar pattern of model choice as in Study 1. Further, this pattern tended to depend on correct trait attribution, holding only for children who correctly identified the model traits.

General Discussion

The present studies investigated the cognitive foundations of preschoolers' selective trust. In the two studies, children did not simply engage in global impression formation as they showed patterns of selective model choice despite the fact that both models showed high (or low) competence. Furthermore, children did not prefer one model in their overall evaluation (i.e., think she was nicer etc.). Nor did they draw narrow inferences on an exclusively behavioral basis, but generalized competences to related behaviors (with the exception of the problem-solving knowledge tasks in Study 2 which were not generalized accordingly). This pattern of generalizations supports the idea that children's selective trust builds upon trait reasoning. Children inferred a trait from shown behavior and chose models as a function of this trait information and the requirements of a given task, preferring, for example, a model who was good at lifting things and who was thus considered as strong for all strength related problems regardless of whether the required behavior matched the demonstrated behavior.

Crucial additional and novel support for the claim that children engage in rational trait-based reasoning comes from the finding that the described pattern of selectivity only holds for children who identified both model traits correctly, whereas children who erred in their trait ascriptions showed no selectivity in any type of task in either study (despite successfully identifying the models' competences in the comprehension questions). To our knowledge, this is the first piece of direct evidence that trait-ascription and selective social learning are related on the level of individual performance. These findings strongly speak in favor of the general claim that selective learning is based on rational, trait-based reasoning; and they constitute novel evidence for the more specific claim by Sobel and Kushnir (2013) that selective learning is a rational process if and insofar as the child has the requisite conceptual background knowledge. So far, interpreting some findings (e.g., children generalize attribute X rationally, but attribute Y in halo-like ways) as evidence for this claim was built on indirect plausibility assumptions (e.g., that children have sufficient conceptual knowledge about X, but not about Y) that were not themselves tested empirically within the same design. The logic of the present study, in contrast, rather than relying on plausibility assumptions, allowed us to apply independent criteria for such conceptual competence (in the form of the explicit trait questions) and thus supplies much stronger evidence for Sobel and Kushnir's (2013) framework.

A potential fundamental concern with the present studies, though, might be that for some of the tasks that were used here it is actually not clear whether they have a unique rational solution. Both when answering the trait questions ("Who is smarter?"), and in the model choice tasks ("Who would be better at performing this labeling/problem-solving knowledge task?"), in a strict sense children do not have enough information to answer with certainty: They have seen positive evidence that Model 1 was smarter than average, but no evidence in this respect concerning Model 2 at all. So, in principle, there might be two kinds of rational strategies: A conservative and skeptical rational strategy aiming at perfect certainty would be to reject the question and demand more information ("I do not know for certain who is smarter or better at this task. I need more information"). But in a sense, this is true for all defeasible, inductive inferences that are made with less than perfect certainty: Even after meeting the millionth white swan, one could reply to the question what the next swan will look like conservatively with "I am not sure, I need more information." This is why, from a theoretical point of view, the problem of induction is a problem after all. And this is why, from an empirical point of view, studies on inductive inferences, for example category-based induction tasks usually use forced-choice tasks to encourage children to make uncertain inferences (e.g., Gelman & Markman, 1986). However, a more liberal rational strategy is to engage in inductive inferences under uncertainty, reasoning along the following lines: "The accurate labeler is definitely remarkably smart, and in the absence of information to the contrary, the good lifter is probably average in terms of smartness, therefore it is likely that the accurate labeler is smarter and better at the labeling/problem-solving knowledge task." Like studies on inductive inferences in general, and like studies on selective trust in particular, we used forced-choice measures to elicit such more liberal rather than conservative rational answers.

Now, it is true that null results in the present context would have been very difficult to interpret: On the one hand, they might have reflected global impression formation and thus the lack of trait ascription and selective trust. But on the other hand, children might have been perfectly capable of trait ascription and selective trust in principle, yet simply preferred conservative rather than liberal rational strategies. Fortunately, however, no such ambiguity arises in the case of the positive results found here: the response patterns of the children answering trait and model choice questions correctly clearly suggest that they did engage in trait ascription and selective trust and made use of the liberal, inductive rational strategy.

How do these early forms of trait ascription found here relate to the divergent findings of much later emergence of trait inferences in the corresponding literature? After all, we did not use any of the measures previously shown to make trait ascription tasks significantly easier like measuring the anticipation of emotions or mental states (e.g., Heyman & Gelman, 1999) or splitting the chain of inferences to reduce cognitive demands (see Liu et al., 2007). However it is possible that inferences in the present studies were still easier because in the demonstration phase children saw explicit samples of behavior instead of hearing abstract stories about a character. A second possible explanation lies within our rather broad definition of traits. In much of the literature on the development of trait ascription, traits are narrowly conceptualized as internal, psychological characteristics that are stable over time and situations and have causal influence on behavior (see Rholes et al., 1990). However, on a broader reading used in much other work, traits are understood simply as nontransient psychological attributes of persons that allow inductive inferences without a much more specific commitment as to the internal structure or stability of the attributes (e.g., Lockhart, Chang, & Story, 2002). And it is clearly this wider reading that underlies the approach of the present study. An open question for future research is thus whether these differently narrow or wide concepts of "traits" might explain some of the divergence in findings.

Regarding the question whether the same cognitive processes are in play when dealing with positive (competence) and negative (incompetence) information, the results showed analogous patterns of inference in the two cases, yet the results in the case of low-competence models (Study 2) appear to be somewhat less clear, as children did not choose selectively in the problem-solving knowledge tasks. Why was this so? One possibility is that this reflects a negativity bias (Vaish, Grossmann, & Woodward, 2008): The greater salience of negative information could have caused more children to generalize below-average competences more broadly and thereby engage in global impression formation in Study 2. In accordance with this explanation, in Koenig and Jaswal's (2011) study, children showed global impression formation for low-competence models but not for high-competence models (pitchfork effect). The less clear results we obtained for the low-competence models (Study 2) might thus reflect a developmental shift between strategies (from global impression formation to trait reasoning) that occurs somewhat later for negative than for positive characteristics. However, because children in fact did show selective model choice in Study 2, this explanation is unlikely to tell the whole story.

A second, complementary possibility is that the differences in the inference structures of the tasks might have been crucial. In Study 1, reasoning was quite straightforward with only two inferences involved in order to arrive at the rational answer: (a) which trait is required and (b) who is the model scoring high on that trait. In Study 2, however, three inferences were required: (a) which trait is required, (b) who is the model scoring low on that trait therefore to be avoided, and (c) by exclusion, which model should be opted for instead. This additional step of reasoning makes the inferences in Study 2 structurally more complex. Thus, in Study 2 some children who competently inferred model traits might still have failed to engage in the last two steps. Such an explanation is in line with findings in other areas such as theory of mind (Friedman & Leslie, 2004) or counterfactual reasoning (German & Nichols, 2003) showing that inferential complexity measured by the number of inferential steps needed directly affects children's performance. It remains a future challenge to design a study with low- and high-competence conditions that build upon similarly complex steps of reasoning.

Concerning selective trust, how do the present results relate to previous findings? First of all, the present studies partly replicate, partly extend the findings of a closely related previous study (Fusaro et al., 2011). Fusaro and colleagues, following similar research questions, presented children with either an accurate labeler, demonstrating basic competence, and an inaccurate labeler (accuracy condition) or a successful and an unsuccessful lifter (strength condition). In the test phase, children decided which model labeled a novel object correctly, who lifted an object, and they answered questions about behavioral predictions in different domains and judged the models on three traits. The results in the accuracy condition indicated that children preferred the accurate over the inaccurate model for the labeling event, the labeling-related behavioral prediction and the trait question "smart" only. Our findings are in line with these results and, due to the fine grained variations of the problem solving tasks, add to the distinction between behavioral matching and trait reasoning and expanded the identified pattern to above-average models (Study 1). In the strength condition, Fusaro et al. (2011) found children to draw broad generalizations, choosing the strong over the weak model in about all test questions. Such a pattern was not found in the current studies where the generalizations from strength and knowledge were similar in breadth. Why then this divergence of findings? Probably and plausibly, this is due to the different ways of introducing the models on the strength-dimension. In the present study, the strong versus weak models were introduced as a minimal contrast pair such that they only differed in the crucial respect whether or not they were able to lift objects. In Fusaro et al. (2011), in contrast, the two models differed in their ability to lift, but also in the accuracy of their announcements because both uttered "I will lift this" before successfully/unsuccessfully trying to lift. Arguably, thus, the weak model was not only weak but also inaccurate (in estimating her own capacities), which, in fact, would make broader inferences concerning the contrast between the two models perfectly rational.

More generally, how can the present results showing trait-based rational selective trust be related to the many previous findings suggesting much simpler processes underlying selective learning in terms of global impression formation and the like? Possibly, this seeming inconsistency is due to the fact that both processes are within the repertoire of preschoolers' cognitive strategies but are selectively engaged in as a function of the structures and demands of certain tasks and problems. Children might be cognitively parsimonious, reverting as a default to simple heuristic strategies (global impression formation) as long as they yield determinate answers (e.g., when one model is in some way better, more competent or likable than the other, as in Bascandziev & Harris, 2014; Brosseau-Liard & Birch, 2010; Cain et al., 1997; Fusaro et al., 2011) and consult more sophisticated strategies (such as trait reasoning) only in situations where the simple strategies do not provide unique solutions (e.g., when models show similar degrees of competence, but in different domains as in the present studies and Corriveau, Kinzler, & Harris, 2013; Kushnir et al., 2013; Landrum et al., 2013, Exp. 1; Lutz & Keil, 2002). Alternatively, children's default might be to engage in rational trait-based selective learning, but they might fall back on simpler strategies when they lack sufficient conceptual background knowledge about the domain in question (Sobel & Kushnir, 2013).

In conclusion, the present studies bring together two fields of research—on the development of selective trust and trait ascription, respectively—that so far were largely unrelated. The findings reported here suggest that under suitable circumstances, 4- and 5-year-olds' selective model choice is based on rational inferences rather than global impression formation or behavior matching, and in fact these inferences are based on the ascription of relevant traits.

References

Bascandziev, I., & Harris, P. L. (2014). In beauty we trust: Children prefer information from more attractive informants. *British Journal of Developmental Psychology, 32,* 94–99. http://dx.doi.org/10.1111/bjdp.12022

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. http://dx.doi.org/10.1111/j.1467-7687.2009.00932.x

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. http://dx.doi.org/10.1111/j.1467-9507.1997.tb00094.x

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science, 12,* 426–437. http://dx.doi.org/10.1111/j.1467-7687.2008.00792.x

Corriveau, K. H., Kinzler, K. D., & Harris, P. L. (2013). Accuracy trumps accent in children's endorsement of object labels. *Developmental Psychology, 49,* 470–479. http://dx.doi.org/10.1037/a0030604

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology, 27,* 331–342. http://dx.doi.org/10.1348/026151008X310229

Friedman, O., & Leslie, A. M. (2004). Mechanisms of belief-desire reasoning: Inhibition and bias. *Psychological Science, 15,* 547–552. http://dx.doi.org/10.1111/j.0956-7976.2004.00717.x

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology, 110,* 561–574. http://dx.doi.org/10.1016/j.jecp.2011.06.008

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183–209. http://dx.doi.org/10.1016/0010-0277(86)90034-X

German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6,* 514–523. http://dx.doi.org/10.1111/1467-7687.00309

Harris, P. L. (2007). Trust. *Developmental Science, 10,* 135–138. http://dx.doi.org/10.1111/j.1467-7687.2007.00575.x

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA: Belknap Press of Harvard University Press. http://dx.doi.org/10.4159/harvard.9780674065192

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 366,* 1179–1187. http://dx.doi.org/10.1098/rstb.2010.0321

Heyman, G. D. (2009). Children's reasoning about traits. *Advances in Child Development and Behavior, 37,* 105–143. http://dx.doi.org/10.1016/S0065-2407(09)03703-3

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child Development, 70,* 604–619. http://dx.doi.org/10.1111/1467-8624.00044

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14,* 106–111. http://dx.doi.org/10.1111/j.1467-7687.2010.00965.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261–1277. http://dx.doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development, 82,* 1634–1647. http://dx.doi.org/10.1111/j.1467-8624.2011.01618.x

Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?": The distinction between causal knowledge and word knowledge guides preschoolers' selective requests for information. *Developmental Psychology, 49,* 446–453. http://dx.doi.org/10.1037/a0031649

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science, 16,* 622–638. http://dx.doi.org/10.1111/desc.12059

Li, Q.-G., Heyman, G. D., Xu, F., & Lee, K. (2014). Young children's use of honesty as a basis for selective trust. *Journal of Experimental Child Psychology, 117,* 59–72. http://dx.doi.org/10.1016/j.jecp.2013.09.002

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. http://dx.doi.org/10.1111/j.1467-8624.2007.01082.x

Lockhart, K. L., Chang, B., & Story, T. (2002). Young children's beliefs about the stability of traits: Protective optimism? *Child Development, 73,* 1408–1430. http://dx.doi.org/10.1111/1467-8624.00480

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73,* 1073–1084. http://dx.doi.org/10.1111/1467-8624.00458

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology, 49,* 404–418. http://dx.doi.org/10.1037/a0029500

Over, H., Carpenter, M., Spears, R., & Gattis, M. (2013). Children selectively trust individuals who have imitated them. *Social Development, 22,* 215–224. http://dx.doi.org/10.1111/sode.12020

Rholes, W. S., Newman, L. S., & Ruble, D. N. (1990). Understanding self and other: Developmental and motivational aspects of perceiving persons in terms of invariant dispositions. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 369–407). New York, NY: Guilford Press.

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55,* 550–560. http://dx.doi.org/10.2307/1129966

Robinson, E. J., & Einav, S. (Eds.). (2014). *Trust and skepticism: Children's selective learning from testimony*. New York, NY: Psychology Press.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72,* 1054–1070. http://dx.doi.org/10.1111/1467-8624.00334

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review, 120,* 779–797. http://dx.doi.org/10.1037/a0034191

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134,* 383–403. http://dx.doi.org/10.1037/0033-2909.134.3.383

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development, 18,* 61–71. http://dx.doi.org/10.1002/icd.591

(*Appendices follow*)

Appendix A

Objects and Labels Used in the Accuracy Demonstrations in Study 1 and Study 2

Phase	Common label	Inaccurate label	Expert label
Demonstration	Glasses (Brille)	Magnifier (Lupe)	Varifocals (Gleitsicht*brille*)
	Bike (Fahrrad)	Buggy (Kinderwagen)	Dutch bike (Holland*fahrrad*)
	Airplane (Flugzeug)	Helicopter (Hubschrauber)	Supersonic airplane (Überschall*flugzeug*)
	Sheep (Schaf)	Horse (Pferd)	Coarse-wool Country sheep (Rauwolliges Land*schaf*)
Optional extension after demonstration	Apple (Apfel)	Peach (Pfirsich)	Boscoop apple (Boscoop-*Apfel*)
Optional extension after model rotation	Flower (Blume)	Tree (Baum)	Shrub flower (Strauch*blume*)
Demonstration reminder (between test blocks)	Hammer (Hammer)	Pliers (Zange)	Sledge hammer (Vorschlag*hammer*)
	Noodles (Nudeln)	Rice (Reis)	Fussili noodles (Fussili-*Nudeln*)
Optional extension after demonstration reminder	Hat (Hut)	Helmet (Helm)	Panama hat (Panama-*Hut*)

Note. Original labels in German language are reported in parentheses.

Appendix B

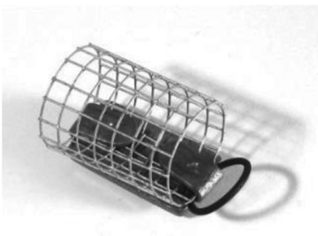
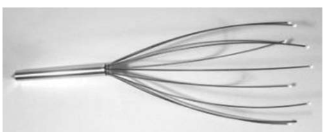
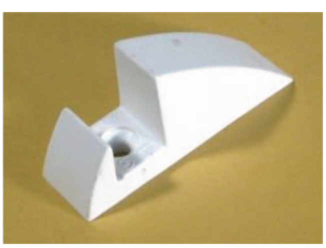
Objects Used in the Strength Demonstrations

Phase	Heavy object (lifted by strong model)	Light object (not lifted by weak model)
Demonstration	Potato sack	Cup
	Big suitcase	Storybook
	Fully-stacked beverage crate	Roll of toilet paper
	Big metal toolbox	Small pepper mill
Optional extension after demonstration	Potato sack	Cup
Optional extension after model rotation	Big bucket with water	Hole puncher
Demonstration reminder (between test blocks)	Huge pile of books	Small football
	Big pot with water	Bunch of keys
Optional extension after demonstration reminder	Big metal toolbox	Small pepper mill

(*Appendices continue*)

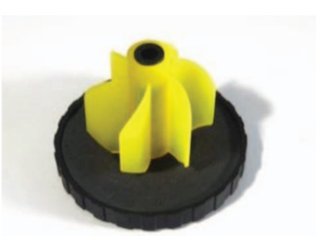
Appendix C

Objects, Explanations, and Test Questions Used in the Test Blocks

Objects by task blocks	Introduction/explanation	Test question
Knowledge task block		
	"Look what they have here. Do you know this?" If they guessed sth.: "I don't think it's a . . ." "Let's see if [Ms. Green] and [Ms. Red] know what this is."	e.g., "[Ms. Green] said that's a [Wugg] and [Ms. Red] said that's a [Flep] What do you think this is?" (novel labels used: Wugg, Flep)
	See above	See above (novel labels used: Blicket, Doso)
	See above	See above (novel labels used: Mido, Toma)
 (spare object)	See above	See above (novel labels used: Dano, Gopi)
Strength task block		
	Instruction: "Look, there are four mice. All of them want to go into the mouse hole. If you hold the blue part and push the red part, the mice move."	"Who can manage to push all the mice into the mouse hole?"
	Instruction: "Look, they have a heavy stone. Would you like to try and lift it?"	"They need to lift it and carry it a long way. Who can do that?"
	"You can move this by squeezing together the two blue parts."	"Who can squeeze these things very far together?"

(*Appendices continue*)

Appendix C (*continued*)

Objects by task blocks	Introduction/explanation	Test question
Problem solving strength task block		
	"Look, they have this rope. But the problem is there's a big knot. They want to remove the knot because they want to play skipping rope. Would you like to try and remove the knot?"	"Who would be able to open the knot?"
	"Look, there's a mouse and there's a piece of cheese. The mouse wants to go to the cheese because she wants to eat it. If you hold the green part and pull and move the brown part, the mouse moves."	"Who can bring the mouse to the cheese?"
Problem solving neutral task block		
	"Look there's a puzzle with some half animals on each card. If you put them together, you can form full animals. Here's a tiger for example. You can make a square with all animals being put together correctly."	"Who can put them together with all animals being correct?"
	"Look, they have this cube. But if you pull one edge, it's no longer a cube. Would you like to try to put it together again?"	"Who can put it together again?"
Problem solving knowledge task block		
	"Look what they have here. Do you know this?" If they guessed sth.: "I don't think it's a . . ."	"Who knows what this is good for, what you can do with it?"
	See above	See above
 (spare object)	See above	See above

See the online article for the color version of this appendix.

(*Appendices continue*)

Appendix D

Amounts of Children Who Answered Comprehension Questions Incorrectly and Needed Extended Demonstrations or Were Excluded From Analyses

	Study 1: Comprehension questions initially incorrect		Study 2: comprehension questions initially incorrect	
Test phase	Correct answers to comprehension questions after extension	Excluded from analysis because answer still incorrect after extension	Correct answers to comprehension questions after extension	Excluded from analysis because answer still incorrect after extension
Accuracy demonstration	2	1	2	0
Strength demonstration	0	0	1	1
Model rotation	7	2	9	3
Demonstration reminder (between test blocks)	4	5	3	4

Received October 21, 2014
Revision received June 8, 2015
Accepted July 15, 2015 ■