The basis of epistemic trust:

Reliable testimony or reliable sources?

Melissa A. Koenig

Institute of Child Development

University of Minnesota

&

Paul L. Harris

Graduate School of Education

Harvard University

Abstract

What is the nature of children's trust in testimony? Is it based primarily on evidential correlations between statements and facts, as stated by Hume, or does it derive from an interest in the trustworthiness of particular speakers? In this essay, we explore these questions in an effort to understand the development course and cognitive bases of children's extensive reliance on testimony. Recent work shows that, from an early age, children monitor the reliability of particular informants, differentiate between those who make true and false claims, and keep that differential accuracy in mind when evaluating new information from these people. We argue that this selective trust is likely to involve the mentalistic appraisal of speakers rather than surface generalizations of their behavior. Finally, we review the significance of children's deference to adult authority on issues of naming and categorization. In addition to challenging a purely inductive account of trust, these and other findings reflect a potentially rich set of tools brought by children to the task of learning from people's testimony.

According to Hume, we trust in others' testimony because we have experienced a long series of correspondences or veridical associations between testimony and reality. His proposal is that because we have learned from past experience that testimony usually corresponds with the facts, we can, via induction, trust it on those occasions when the relevant facts cannot be checked (Hume, 1739/1977).

This elegant portrait has empirical implications for the development of trust. We will focus on two facets of this view, one that receives empirical support and one that does not. The central facet that does receive support is that children are sensitive to whether informants have proved accurate in the past and this affects their subsequent trust in that informant. The facet of Hume's view that does not receive support is the scope of the argument. Children may not be filtering their general experience of language to arrive at an expectation that testimony is reliable in some general sense. Rather, they attend to characteristics of speakers and form judgments about the trustworthiness of particular people. Children's surprisingly careful attention to the individual source of a given claim highlights the importance of social cognition in early trust. In addition, for Hume, a global attention to linguistic correspondences paves the way toward a generalized credulity when, in fact, developmental evidence demonstrates that such a generalized credulity is not true of young children, even infants.

We review recent empirical literature supporting these conclusions. We begin by describing experiments showing that infants' attention to familiar labeling events goes beyond sensitivity to the accuracy of those labels. It also includes an awareness of the role of a speaker's perception and perhaps, her state of knowledge. Next, we discuss recent findings with 24-month-olds revealing an early ability to limit credulity toward

previously inaccurate sources. We then consider older children's ability to trust the more accurate of two informants and discuss how well children can differentiate between informants when each has been inaccurate, but to varying degrees. We further challenge the Humean account by arguing that children's judgments of trustworthiness likely reflect mentalistic appraisals of informants rather than empirical generalizations about reliability. Finally, we discuss evidence for a trust that is based not on correspondences or positive evidence but instead is extended 'on faith'. Here, we review some of the factors that delineate trustworthy sources for children while discussing children's trust in authoritative figures on matters of naming and categorization.

Infants. The first step to limiting one's credulity is to discriminate between messages that conflict versus concur with prior knowledge. To check for the early emergence of this discrimination, infants aged 16-month-old infants were presented with pictures of familiar objects and heard either labels that matched the objects ('True condition') or labels that did not match the objects ('False condition') (Koenig & Echols, 2003). The source of the utterance was also varied. Across a series of studies, labels were produced by either a human speaker who labeled while looking at the objects, a visible audio speaker, a concealed audio speaker, or a human speaker with her back toward the objects. We recorded how long infants looked at the object and the labeling source. Consistent with Hume's view, if infants hold a general expectation that labels match their corresponding referents, then we would expect to see a difference between the True and False conditions, independently of who or what produces the labels. However, if infants' attention to accuracy is marked by an interest in the source of the

message, then infants' behavior should be mediated not by accuracy alone but also by the nature of the labeling source.

Figure 1. Infants' total looking time to the labeling source during true and false labeling in Studies 1 and 2.
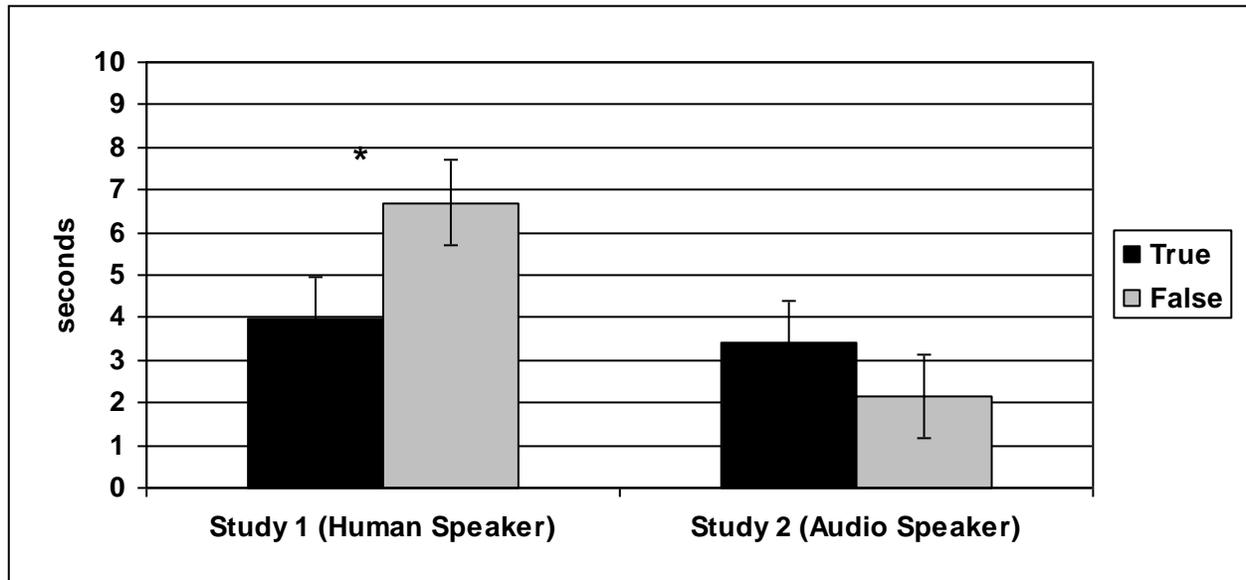

p<.05

In Study 1, infants looked longer toward the human speaker when the speaker produced false rather than true labels. However, in Study 2, when the source of the utterance was inanimate, infants did not look longer at the source during false versus true labeling. Thus, this comparison provides evidence that infants give marked attention to a source of false labels – but only when it is human. Even more compelling evidence that children give marked attention to false labels from human speakers comes from infants' own spontaneous productions. As Figure 2 makes clear, infants commented on the labeling events by accurately labeling the object they saw. This means that in the False condition, infants' own accurate labels often came on the heels of the speaker's false labels, suggesting that they attempted to correct the speaker. Strikingly, infants were

more likely to produce such 'corrections' when false labels came from human speakers rather than an inanimate source. In fact, 15 out of 16 infants tried at least once to correct the speaker. Such results are in line with Pea (1982) who found that 18- to 24-month-old infants explicitly denied false affirmations (e.g., "That's a car" in reference to a ball) by saying "no".

Figure 2. Infants' spontaneous productions. Number of infants (of 16) stating the correct name.
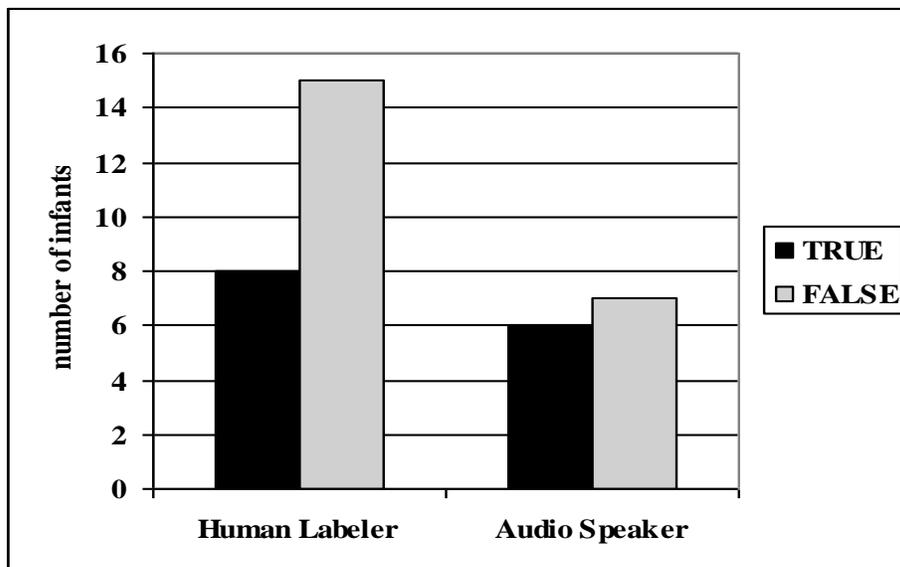

Across both Studies 1 and 2 above, infants looked longer at the object during true labeling than during false labeling. This finding corresponds with research on lexical development suggesting that infants prefer that labels match the object with which they have been associated (Hirsh-Pasek & Golinkoff, 1996; Schafer & Plunkett, 1998). This well-replicated pattern is exactly what Hume would predict. In principle, the finding that infants were surprised by false labeling from a human source could be seen as further support for Hume's reductionist account. After all, it may simply be that infants' experience has taught them to associate a certain type of behavior – labeling – with

entities observed to have that capacity – humans (and *not* audio speakers). People, not radios, tend to talk about things that are relevant and consistent with infants' experience. Thus, infants' expectations for truth may be violated whenever there is a mismatch between a referent and the label supplied by a human speaker.

We tested this possibility by presenting infants with a true or a false human labeler who faced backward and gazed away from the objects (Koenig & Echols, 2003, Study 4). If infants' attention to false labelers is simply due to a violated association between humans and accurate labeling, we would expect infants to look longer to the false labeling source regardless of her orientation. However, if the results of Study 1 were due to an early intentional understanding, we would expect to see a difference between infants' behavior toward labelers who lack versus possess perceptual access to objects. As illustrated by Figure 3, infants did not seem particularly surprised when a speaker produced false labels while gazing away from objects. If anything, they appeared surprised to hear true labels produced by someone with their back turned toward the objects.

Figure 3. Infants' looking time to the human speaker during true and false labeling events of Studies 1 and 4.
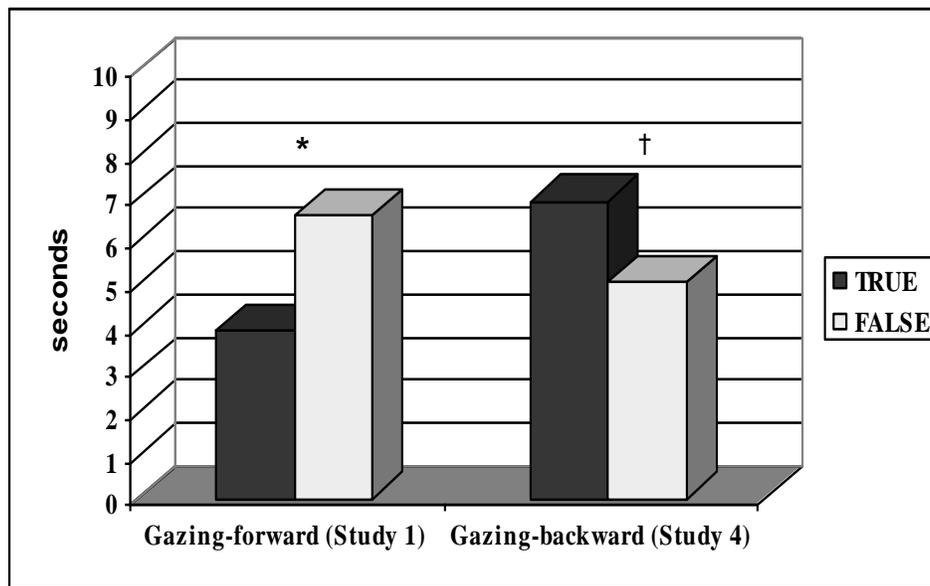
* *p* < .05
† *p* = .065

In sum, infants use their knowledge of word meanings to identify, correct and deny false statements (Koenig & Echols, 2003; Pea, 1982). Second, infants' attention to false human labelers reveals more than an expectation, based on prior associations, that speakers are generally truthful: Infants appear sensitive to the epistemic conditions – conditions marked by a speaker's perceptual access – under which people typically name objects correctly. Thus, infants are not only encoding the truth value of messages. They also attend to characteristics of the speaker; they are surprised by false labels if a speaker has perceptual access and by true labels if a speaker lacks perceptual access.

Toddlers. Infants' ability to recognize, correct and deny assertions that conflict with what they know, however effective, does not tell us whether they evaluate the potential trustworthiness of an informant. For example, we do not know from the studies reported above whether infants were encoding the particular identity of the inaccurate informant. Similarly, the expectation that human speakers label things in accord with

their perceptual experiences leaves open the question of whether infants gauge the truth of an unfamiliar claim. To evaluate infants' capacity for such selective trust, it is crucial to present testimony that they cannot check against their own experience.

In a recent examination of this issue, we presented 24-month-olds with an adult who labeled familiar objects either accurately or inaccurately in a between-subjects design (Koenig & Woodward, under review). Then, to assess infants' capacity for mistrust, the adult provided novel labels for the test objects and infants were asked to select the target. If children unreflectively assent to whatever adults tell them, then 24-month-olds may not evaluate novel labels differently from an accurate versus an inaccurate speaker. Children might accept all claims as true, despite the speaker's prior inaccuracies. On the other hand, if children are open to the possibility of speaker error or ignorance then attention to the accuracy of prior claims may trigger subsequent mistrust.

We also explored a possible source of individual differences in children's mistrust by obtaining a measure of children's productive vocabulary size (Fenson, Pethick, Renda, Cox, Dale, & Reznick, 2000). Recognizing claims as false depends on knowing the relevant word meanings and identifying violations of use. Presumably, as children's vocabulary knowledge increases, so does their ability to identify claims as mistaken. This experience of identifying false claims and their contexts may help infants identify the conditions under which speaker are likely to make errors. On the other hand, the risk of deception and the cost of accepting false information is the same for all learners (Dawkins & Krebs, 1978). Thus, a basic form of caution would be to submit all input to the same checking procedure, regardless of background lexical knowledge. By asking parents to report on their child's productive vocabulary, we could investigate whether

children's mistrust of an inaccurate source was influenced by their prior knowledge of words.

Overall, we found that the accuracy of a speaker's labeling history affected toddlers' evaluation of novel object labels (Figure 4). Toddlers were less likely to indicate the target object in response to a previously inaccurate speaker than in response to an accurate speaker. Consistent with Hume's proposal, even toddlers appeal to their past experience when evaluating novel and thus, newsworthy claims. Inconsistent with Hume's proposal, however, is the lack of evidence for a generalized credulity. If Hume were right, children's past experience with reliable language use should have fostered a similar level of trust across both the accurate and inaccurate conditions. Instead, children's trust and mistrust depends on the reliability of the particular source. Importantly, this raises the possibility that children's trust in the unmarked or undefeated case, is driven not by their general experiences with language but by experiences with particular individuals. Furthermore, while children with high vocabularies performed better than children with low vocabularies overall, there was no indication that vocabulary knowledge mediated children's ability to mistrust the inaccurate source. Thus, regardless of differences in prior word learning, the reliability of the speaker had an effect on children's propensity to accept a new word.

Figure 4. Proportion correct (of 4 total tests) and comparisons against chance in 24-month-old children with both high and low vocabularies (N = 64).

High vocab   Low vocab

*   *Accurate*
    Inaccurate
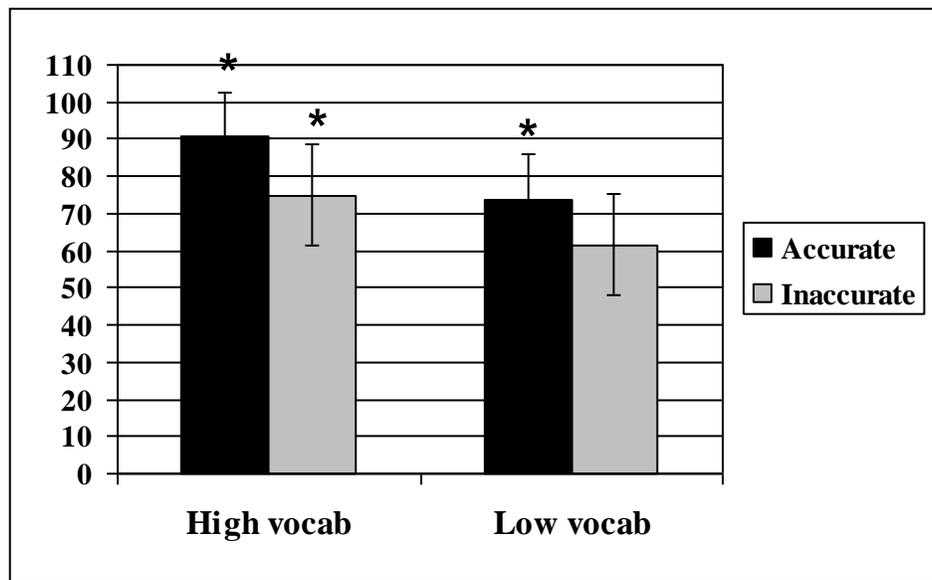
*   *p* < .001

This finding is significant for several reasons. First, it is the first demonstration of

selective trust with children as young as 24 months of age. It is particularly interesting in

light of toddlers' reliance on cues to a speaker's referential intent (e.g., eye gaze,

pointing, naming frame) when determining the meaning of novel words (Akhtar,

Carpenter & Tomasello, 1996; Baldwin, 1991, 1993; Baldwin, Markman, Bill,

Desjardins, Irwin & Tidball, 1996; Baldwin & Moses, 2001; Tomasello & Barton, 1994).

In the study reviewed here, toddlers interacted with a speaker who presented several of

these powerful cues: she referred to an object with a novel name in a typical naming

phrase while pointing and gazing at the object. Toddlers' doubt reveals an ability to

resist the force of these cues having been given signs of a speaker's unreliability.

Second, it is interesting that sensitivity to the source can be found in the language

domain at all. Language is a conventional system in which all speakers (in a community)

use the same words to convey given meanings (Clark, 1993). In principle, infants might

neglect to encode or represent the particular source of their knowledge of any given word meaning because speakers typically share and affirm the same word meanings in their conversation. Thus, it is interesting that children attended to source information rather than simply accepting new words "on faith". Perhaps a speaker who violates shared tenets is particularly salient to a child, especially one who appreciates the shared nature of word knowledge (Graham, Stock & Henderson, 2006; Henderson & Graham, 2005). More specifically, the apparent unreliability of a source may carry special significance for children (and adults) in domains of high agreement or shared consensus like language, religion or folk biology (Harris & Koenig, 2006; Kalish & Sabbagh, 2007).

Third, similar research with older children has involved an accurate and an inaccurate informant simultaneously with children being asked to choose between the two sources. Such research has left it unclear as to whether children would mistrust an inaccurate informant when no alternative source of information was available (Clément, Koenig & Harris; 2004; Jaswal & Neely, 2006; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, Harris, in press). The present results suggest that even toddlers sometimes resist information provided by a single, unreliable informant. This result is noteworthy because, intuitively, it seems harder to reject information from a single authority than to identify the more reliable of two sources. It also implies that children can monitor the quality of information an informant provides and use that to assess her reliability in the course of everyday interactions.

Finally, the finding that young children resist novel information from an inaccurate source begins to test Reid's claims about our a priori trust in testimony. For Reid (1764/1997), credulity is a "gift of nature" and the "strongest in childhood" (p. 197).

On Reid's view, testimony is treated like first-hand experience; we accept what we are told just as we accept the 'testimony of our senses' or the 'testimony of our memory'. However, infants who express a certain degree of doubt toward an utterance can be credited as discriminating between information conveyed in an utterance, and the reality it comments on. In line with Reid's epistemology, the infant could construe the environment as presenting an undifferentiated barrage of information and whenever any information – memorial, testimonial, or perceptual – conflicts with their prior beliefs, they respond by simply updating their beliefs. The finding that, by 24 months of age, linguistic utterances are treated as possibly untrue raises the possibility that infants never treat language this way, as akin to first-hand experience. The understanding that language comments on experience may be there from the beginning of language development (Koenig & Woodward, under review).

Preschoolers. In each of the studies reviewed above, infants were presented with unambiguous evidence of a speaker's unreliability in that the unreliable source was consistently inaccurate across all trials. However, we know that, in the real world, such speakers do not exist. Instead, children confront speakers who are mostly accurate but differ in the extent to which they err or reveal ignorance. As adults, we recognize this mix of accuracy and inaccuracy yet still we judge certain informants to be better sources than others. Recent research by Pasquini et al. (2007) asked whether preschoolers weigh occasional inaccuracy against a backdrop of general reliability. Children were presented with a series of familiar objects (e.g., cup, dog, shoe, ball) and watched as two informants offered conflicting labels in answer to a bystander's inquiry. In a baseline condition, children were presented with an informant who was accurate across four trials and one

informant who was inaccurate across all four trials (100% vs. 0% condition). In a second condition, one informant was consistently accurate whereas the other was inaccurate on three out of four test trials (100% vs. 25% condition). In a third condition, one informant was accurate on three out of four trials whereas the other was consistently inaccurate on all four trials (75% vs. 0% condition). Finally, in a follow-up experiment, children were presented with an even more difficult comparison: one informant was accurate on three out of four trials whereas the other was accurate on one out of four trials (75% vs. 25% condition).

Children's selective trust was examined with the following three types of probe. First, before and after the impending test period, children were asked for an explicit judgment, "Which person was not very good at answering questions (about the names of objects)"? Second, children were presented with a series of novel objects whose names they did not know and they were invited to ask one of the informants what each novel object was called, "I bet one of these people can help. Who would you like to ask?" Third, regardless of whom children had chosen to ask, each informant offered a different name for the novel object and children were invited to choose one of the two supplied names, "What do you think it's called?"

The two age groups studied here, 3- and 4-year-olds, differed in their patterns of response (see Figure 5). Four-year-olds trusted the more accurate informant across all the comparisons tested. They successfully identified the speaker who was "not very good at answering questions", they directed their questions toward the more accurate informant and when the informants offered conflicting novel terms, they endorsed what the more accurate informant said. Three-year-olds, in contrast, trusted the more accurate informant

for two comparisons only: the '100% vs. 0%' condition and the '100% vs. 25%' condition. That is, they could make appropriate identifications, seek help from the accurate source and endorse her novel labels only if they were able to pick out a consistently accurate informant.

Figure 5. Proportion of correct responses and comparisons against chance by age and condition.
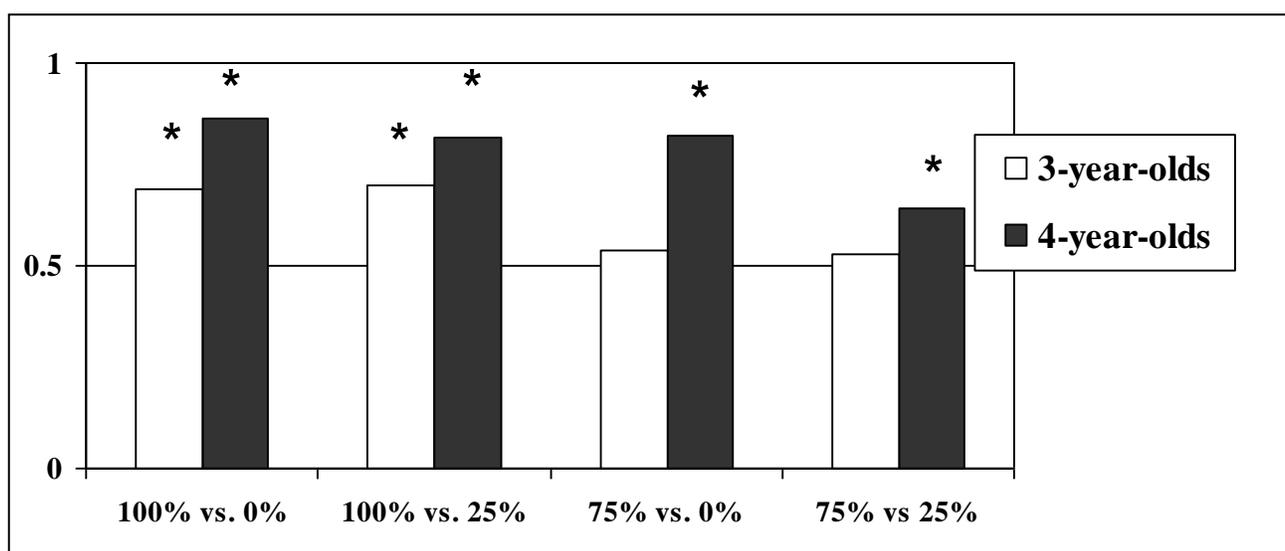
* p < .05

When presented with two informants who were less than consistently accurate, in the 75% vs. 0% and the 75% vs. 25% conditions, 3-year-olds responded in a random fashion. That is, they failed to identify the informant who was "not very good at answering questions", they indiscriminately asked both informants for information and endorsed either informant's novel label. Crucially, in these conditions involving a 75% accurate informant, the more reliable informant made only one error whereas, in the two conditions discussed above involving a 100% accurate informant, the more accurate

informant made no errors. Such a pattern suggests that 3-year-olds may be intolerant of even a single error and treat the person responsible for that error as no more trustworthy than someone who has made many errors.

In addition to this differential pattern of responding, it is worth noting that 4-year-olds performed better than 3-year-olds overall (Pasquini et al., 2007). Such age differences have been found in a host of previous experiments that presented children with consistently accurate and inaccurate informants (Clément et al., 2004, Experiments 1 & 2; Koenig, Clément & Harris, 2004; Koenig & Harris, 2005, Experiment 1). The robust age change between 3- and 4-years on tasks involving accurate as opposed to ignorant informants (see Koenig & Harris, 2005, Experiments 2 and 3) has led to speculation that such differences are based in well-documented differences in mental state reasoning (Clément et al., 2004; Koenig & Harris, 2005; Robinson et al, 1995; Welch-Ross, 1999). Four-year-olds are typically better than 3-year-olds at identifying when someone holds a false belief and at recognizing its implications for what that person will say and do (Wellman, Cross & Watson, 2001). In the studies discussed above, perhaps 3-year-olds were puzzled by the false claims whereas 4-year-olds attributed them to the informant's false beliefs. In that case, 3-year-olds might not expect a speaker's inaccuracy to persist whereas 4-year-olds might plausibly do so.

Results from more recent research now call this conceptual explanation into question. First, Pasquini and colleagues (2007) tested the above hypothesis and found no relation between standard measures of false belief understanding and children's selective trust. Three-year-olds performed above chance in the 100% vs. 0% and the 100% vs. 25% conditions, as noted above, and yet still performed systematically below chance on

the false belief task. Indeed, many children who failed the false belief task went on to display selective trust in the more accurate informant. Thus, we conclude that while an understanding of false belief may yet prove to play a role in selective mistrust, it does not appear to be critical for the mistrust that is provoked by past inaccuracy

Second, 3-year-olds' performance improves with enhanced and simplified versions of the task (Pasquini et al., 2007) and as we discussed earlier, even 2-year-olds' selective trust is evident when they interact with only one informant (Koenig & Woodward, under review). Thus, very young children, even toddlers, are capable of selective trust but may not be as competent or efficient as older children. After all, children's trust likely recruits a set of abilities. First, children must be able to detect when a speaker makes an error. Second, children need to keep track of the speakers' behavior which involves differentiating between them, encoding their separate identities and holding in memory person-specific information about accuracy. Third, children need to use this information to guide future learning from these speakers.

The first requirement – that children recognize mistakes – is in place quite early in development. As discussed earlier, 16-month-infants appear surprised when a person labels objects incorrectly and attempt to correct the speaker (Koenig & Echols, 2003; Pea, 1982).

The second requirement – the encoding, monitoring and recall of source information – is a potential source of variation and development. Indeed, evidence suggests that children's selective trust in the more accurate informant is mediated by their successful monitoring of the inaccurate informant. In Koenig et al., (2004), children were presented with two informants who were consistently accurate and inaccurate

regarding familiar object names. After 3 trials of this type, children were asked for an explicit judgment, "Which one of these people was not very good at naming things?" Then at the end of the study, after receiving no new information about accuracy, children were asked this question again as a check of their memory. On the intervening test trials, each informant provided a distinct novel name for an unfamiliar object (e.g., "That's a dax" vs. "That's a wug") and children were invited to endorse one of them (e.g., "What do you think it's called?"). We assessed children's performance on these test trials by dividing children into two groups according to their ability to correctly monitor the sources and differentiate between them: one group performed perfectly (made correct judgments on 2 out of 2 trials) whereas the other group made errors (see Figure 6).

Figure 6. Proportion of choices directed to the more reliable informant on test trials by age and judgment.
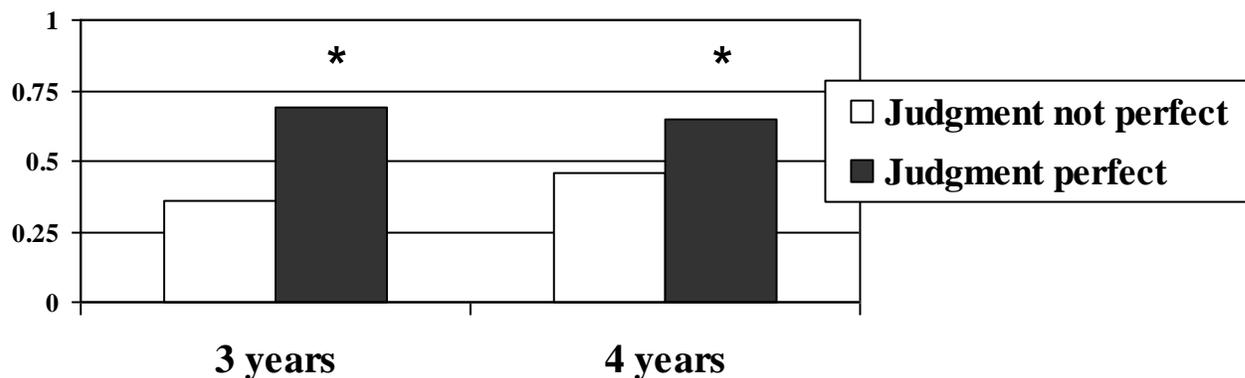
As the above figure illustrates, children in both age groups who successfully differentiated the two informants went on to display selective trust in the accurate informant. In thinking about the age differences reported above, it is important to note

that 50% of the 3-year-olds and 70% of the 4-year-olds performed perfectly on the explicit judgment probe. Thus, age-related improvements in children's ability to trust the more reliable informant are likely due to changes in the ability to differentiate, encode and monitor multiple sources of information. In addition, recent research by Corriveau and Harris (2007) has shown that children who successfully differentiate and monitor the two informants demonstrate trust in an accurate over an inaccurate informant over multiple time points: immediately, after 1 day and after 1 week. Such findings suggest that children's assessments of trustworthiness constitute lasting and enduring judgments. They also bolster the current claim that variation in children's ability to differentiate reliable from unreliable informants will go some way toward explaining their selective trust in a reliable informant – both in the short- and long-term.

The third requirement – the use of source information to make inferences about an informant's future behavior – is another potential source of developmental change and likely hinges, in large part, on the extent to which children treat evidence about reliability as a stable, person-dependant trait (Koenig & Harris, 2005; Miller, 2000). Important differences in trait understanding have been found among children 2 to 10 years of age (Heyman & Dweck, 1998; Rholes & Ruble, 1984; Yuill, 1992).

Of particular relevance here is how to characterize young children's understanding of traits. On one hand, even 2-year-olds have been observed to use trait-like terms such as "nice" and "naughty" (Bartsch & Wellman, 1995; Bretherton & Beeghly, 1982); on the other hand, a child who describes someone as "nice" might be expressing positive feelings or commenting on an action rather than offering a description of what someone is generally like. Thus, one possibility is that young children have a

trait-like conception of reliability but one that is quite superficial. Young children might appreciate that certain people are more reliable sources of information than other sources but they may not see a connection between the reliability and the speaker's mental life. According to this possibility, children treat traits as empirical generalizations regarding overt behavior or outcomes and simply view the inaccurate speaker as a source of 'bad output' and the accurate speaker as a source of 'good output'. In other words, information about a speaker's reliability might be treated like the reliability of faucets or clocks. The other possibility, of course, is that young children view traits, like reliability, as providing a deeper account of a speaker's behavior. On this account, children would not treat traits as descriptive summaries but would understand that the behavioral expression of a trait is mediated by underlying mental states. As discussed by Heyman and Gelman (2000), it is only with this deeper understanding that traits come to be viewed as constructs that allow for a wide range of inferences about people.

What evidence bears on the question of whether young children's trust is based on mentalistic attributions? First, recall that even 16-month-old infants seem to be doing more than merely encoding correspondences, noticing violations and differentiating 'good' from 'bad' output (Koenig & Echols, 2003). More specifically, they looked longer at a human speaker who labeled objects falsely while gazing at the objects whereas they did not look longer at a human speaker who provided incorrect labels for objects she could not see. In fact, infants looked longer at the human speaker when she provided the correct labels for objects she could not see. This suggests that toddlers and young children are not solely attending to output and raises the possibility that even

infants might distinguish between honest mistakes and errors made out of ignorance (see also Onishi & Baillargeon, 2005).

Second, recent research has given children the opportunity to generalize reliability information across different domains of knowledge (Birch, Luca, Frampton, Vauthier & Bloom, 2005; Koenig & Harris, 2005, Experiment 3). In Birch et al. (2005), children were presented with an accurate and an inaccurate source about the function of common objects and then asked whether they would use that information to trust the more accurate source about object names. Four-year-olds, but not 3-year-olds, used accuracy information about functions when deciding which informant to rely upon for information about object labels. Koenig and Harris (2005) asked children to make inferences in the reverse direction – from a history of speaker's word knowledge to a person's credibility regarding object functions. Children were first presented with an ignorant versus a knowledgeable labeler and then tested with novel object functions. Both 3- and 4-year-old children used the speakers' word knowledge when weighing information about novel object functions. Taken together, these studies reveal that young children's extensions of trust are not based on narrow notions of reliability. That is, children are not simply making surface generalizations about speakers. The fact that children can use past reliability with one kind of information to make predictions regarding another kind of information, suggests that children are not tied to the particular type of reliability they have observed but may be assessing informants in a deeper fashion – as generally more or less knowledgeable (see also Danovich & Keil, 2004; Lutz & Keil, 2002) for children's developing conceptions of expertise).

Finally, if children were working with a superficial concept of reliability, they might be expected to link accurate or knowledgeable behavior with a certain category of people, namely adults. That is, children's judgments might not be about underlying knowledgeability but about who has typically made accurate reports in the past. After all, children appreciate that adults tend to know more than they do (Taylor, Cartwright, & Bowden, 1991), and generally more than other children (Ceci, Ross & Toglia, 1987; Lampinen & Smith, 1995). So, perhaps children's real-world decisions often adhere to a simple rule: 'Trust the adults'. Two recent experiments cast doubt on this hypothesis. First, in research by Robinson and colleagues, preschoolers evaluated a claim by determining whether the speaker had appropriate access to the facts that they were discussing (Robinson, Mitchell & Nye, 1995; Robinson & Whitcombe, 2003). For example, Robinson and Whitcombe (2003) presented 3- to 5-year-old children with conflicting pieces of information and asked them to decide which should be believed. In one task, the child experienced an object directly (e.g., by seeing it) and said that it was red. An experimenter experienced the object in a different modality (e.g., by feeling it) and said that it was blue. A puppet on the sidelines, who appeared puzzled, asked the child, "Which is it, red or blue?" If children adhered to a simple maxim that privileged information from adults, children would be predicted to systematically update their belief to agree with the experimenter. Interestingly, however, children made informed decisions about when to believe contradictory testimony: they updated their original statement when the experimenter had better evidence (e.g., the experimenter had seen the object's color) and they stood by their original claim when their evidence was better than

the experimenter's (e.g., the experimenter had only touched the object). Thus, children weigh a speaker's claims against their own first-hand observation

Note that in some respects, these results are surprising. Children typically perform poorly in classic source monitoring tasks where they are asked to explicitly recall the experience that led to their knowledge (e.g., "How did you know? Did you see it, hear about it or feel it?") (Gopnik & Graf, 1988; O'Neill, Astington & Flavell, 1992; Taylor, Esbensen & Bennett, 1994; Wimmer, Hogrefe & Sodian, 1988). The results discussed above raise the possibility that real-life source monitoring might not involve or require such explicit, descriptive judgments that distinguish sight from inference or touching from listening. Instead, it might be sufficient to judge whether information came from a more (or less) reliable source. In other words, more important than the descriptive details of how you know something may be whether the belief was formed with reliable underlying reasons. In line with these considerations, 4- and 5-year-olds' performance reached ceiling levels when asked questions about the 'best way' to acquire a certain piece of knowledge, in contrast with lower performance on how they actually did so (O'Neill, Astington & Flavell, 1992).

A second reason to doubt that children attribute the trait of reliability to all adults under all conditions comes from research by Jaswal and Neely (2006) who pitted past reliability against the age of an informant. They found that when given a choice to learn new words from either a child or an adult – both of whom had a history of being accurate – children tended to choose the words provided by the adult. When both informants were unreliable in the inaccuracy trials, children trusted neither informant systematically, and often said that neither of the novel names they supplied was correct. However, when

given a choice to learn new words from either a child who had a history of being accurate or an adult who had a history of being inaccurate, children preferred to learn the words provided by the child. Adults' assumption that an interlocutor is truthful is a conversational default that can be overridden. Similarly, children's trust in an otherwise reliable speaker can be defeated by his or her prior inaccuracy.

Cumulatively, these three lines of evidence – infants' sensitivity to human mislabeling, children's extrapolation from one type of information to another, and their sensitivity to a speaker's individual history of accuracy rather than his or her membership in a given social category – suggest that children adopt a mentalistic stance toward the testimony of others. At this point in the research program, it is possible to raise objections to each finding. Arguably, children's extrapolation across information domains might simply index a very global and undifferentiated notion of information; and children might group speakers into epistemic but ultimately non-mental categories such as 'those with appropriate evidence.' We anticipate, however, that future research will show that although these objections can be raised in a local, piecemeal fashion, a coherent account of children's selective trust must appeal to their mentalistic insights.

In sum, Hume's view that trust in testimony is built out of observations of the constant and regular conjunction between what people report and the way the world is leaves out of the picture a crucial component: social cognition. The evidence reviewed increasingly suggests that children are not guided by a generalized credulity based on induction alone but by an awareness that speakers are responsible for utterances, both true and false. By attending to characteristics of the speaker, such as her past reliability or her access to information, children extend trust to certain speakers and not others.

This review is not meant to suggest that young children rarely trust others' testimony or that they treat what they hear with skepticism. On the contrary, in spite of the fact that young children spend much of their time investing trust – quite deservedly – in what reliable informants tell them, this research reveals that children possess an active interest and acumen in evaluating the validity of what they are told.

'Blind' trust. The Humean view of testimony, as discussed above, places a significant burden on prior communication since it implies that epistemic trust is built out of past experiences with claims whose correspondence to facts can be checked. However, in everyday life, children (and adults) often lack such relevant experience with particular informants. Nevertheless, they ordinarily take people at their word. When a speaker reports some fact, the listener believes the report simply because of the trust they place, not in the utterance, but in the speaker. The speaker, in these cases, is the object of appraisal and it is based on her authority that testimony is accepted.

Infants' readiness to learn words from an unfamiliar experimenter, a common finding in the word learning literature, provides indirect evidence that infants ordinarily assume that a speaker can be trusted to provide accurate information. In most word learning experiments, a person previously unknown to the child presents a new object in a laboratory setting, labels it with an unknown name and infants show no signs of mistrusting the speaker but instead, seem to accept and readily learn the label (Baldwin, 1993; Bloom, 2000; Golinkoff, Mervis & Hirsh-Pasek, 1995; Markman, 1989; Woodward, Markman & Fitzsimmons, 1994). Given that children typically have no opportunity to assess a particular experimenter's prior record of accuracy or sincerity,

their willingness to learn new words from an unfamiliar person suggests an a priori openness to the claims of others.

When obvious misstatements occur, infants' rejection stands on firm ground because what they see is in direct conflict with what they hear (Koenig & Echols, 2003; Pea, 1982). As ambiguity is introduced, however, children show more willingness to set aside their own perceptual inferences in favor of an adult's claim. For example, in classic research by Gelman and Markman (1986), 4-year-olds were shown a picture of a fish and told that, "this fish stays underwater to breathe". They were also shown a picture of a dolphin and told that, "this dolphin pops above the water to breathe". When shown an ambiguous object – a dolphin-like shark – and asked what "this fish" would do, children inferred that it breathed underwater, based only on the label the adult provided. Not only do people's words invite children to form categories (Markman & Hutchinson, 1984; Waxman & Markow, 1995; Welder & Graham, 2001), they can lead them to abandon their original classification in favor of another, known category (Gelman & Coley, 1990; Gelman, Collman, & Maccoby, 1986; Gelman & Markman, 1986, 1987; Sloutsky & Fisher, 2004).

More recent research by Jaswal and Markman (2007) has found that even 24-month-olds will give up a convincing, perceptually-based classification in favor of an unexpected one based on a speaker's label. Toddlers were presented with ambiguous hybrid entities (e.g., a cat-like dog or key-like spoon). When asked to make a prediction without any label information, "Can you show me what this one does?", children based their judgments on appearance (e.g., that the cat drinks milk and the key starts a car). However, when told that the same dog-like cat was in fact a dog and asked, "Can you

show me what this dog does?", 2-year-olds overwhelmingly accepted this unexpected label and inferred that the ambiguous creature would eat bones, like a dog. Thus, even in the early stages of categorization, children defer to expert authority (Neisser, 1987). Indeed, the understanding that objects can fall into more than one category and be called by multiple labels (Rex, poodle, pet, dog, animal, etc.) may rest, in part, on children's acceptance of an 'Authority Principle' (Mervis, Pani & Pani, 2003).

Children's willingness to make an inference on the basis of an adult's label, especially when the label conflicts with their own classification, is significant for several reasons. First, it suggests that even very young children appreciate that the 'testimony of their senses' can be misleading. Even though something looks distinctly like a cat, that doesn't necessarily mean it is a cat – it could be a dog. Second, children defer to the testimony of an unfamiliar speaker who offers no explanation for his or her counter-intuitive testimony. In matters of naming and categorization, children often take informants at their word.

This is not meant to suggest that blind credulity drives these decisions. Children do use evidence of reliability when it is available. If possible, children make judgments about the epistemic status of the speaker. For instance, if the speaker shows signs of uncertainty, children are less likely to accept testimony. Four-year-olds are sensitive to explicit linguistic markers that modulate assertions (e.g., 'know' versus 'think' and 'guess') (Moore, Bryant, & Furrow, 1989; Moore & Davidge, 1989). Three-year-olds are less likely to trust counter-intuitive testimony if the speaker has made a mistake and seems distracted or says, "I *think* this is a dog" rather than just labelling the object ostensively (e.g., "This is a dog") (Jaswal & Malone, in press). Regarding even more

subtle cues, children of this age may prove cautious in learning a new word if the speaker hesitates before offering a label (Sabbagh & Baldwin, 2001).

In practice, children are given opportunities to use these cues by parents who are more, or often less, knowledgeable about objects and their names. Sabbagh and colleagues (2007) asked parents to label objects for their 2-, 3-, and 4-year-old children as they would normally do so at home. Objects fell into different classes: a) both parent and child knew the correct label, b) only the parent knew the correct label, and c) neither parent nor child knew the correct label. Parents have various options when they are ignorant about object labels – they can avoid them altogether, they can make up random names to support the local interaction or they can offer labels while marking their own ignorance in some clear way. Preliminary evidence suggests that parents change their strategies over time. For younger children, they avoided labeling objects if they did not know the conventional name. Over time, however, parents provided labels for objects whose names they did not know (i.e., they guessed) but they explicitly marked their guesses – they use verbal, non-verbal, and paralinguistic expressions of their ignorance (Sabbagh, 2007).

Finally, children are mindful of adults' repairs (Clark, 1993). In one task, 2-, 3- and 4-year-olds were taught a word ("mido") for a set of objects. After children received 10 exposures to the new word, the adult said, "Oh, I made a mistake: these aren't midos, they're blickets," and proceeded to teach the new word ("blicket") for the same objects. When asked to identify all the "midos", even the youngest children refused to select from the test objects, saying "no", "don't know" and not responding. But when asked to identify all of the "blickets", children selected the appropriate referents. Consistent with

research reviewed above, young children are tracking signs of speaker ignorance and interpreting their repairs appropriately (Clark & Grossman, 1988; Sabbagh, Wdowiak & Ottaway, 2003).

To sum, Hume's account of our trust in testimony is that by observing that testimony has corresponded with facts of the world, we come to infer that it will likely correspond in the future. This inductive argument finds some support in the finding that children, as young as 2, expect that individuals who have proved accurate in the past will prove accurate in the future. Indeed, we think that a full account of epistemic trust will ultimately, to some degree, appeal to a listener's past experience. However, a solely inductive account falls short of a complete story. First, the basis of children's trust is misconstrued: children's trust is built not on linguistic correspondences in the abstract but on their attention to the particular speakers who often, but not always, produce them. Indeed, as fundamental research by Baldwin has shown, infants learn words not by associating words with temporally contiguous facts but by attending to the eyes of the speaker (Baldwin, 1993; Baldwin et al., 1996; Baldwin & Moses, 2001). Second, it is important to note that when children's decisions about who or when to trust are guided by mental characteristics of speakers, these are features that go beyond simple observational learning. Finally, as discussed above, even when children's trust is blind – even when children place an a priori trust in a speaker – this trust is not 'unreflective'. It too is responsive to a speaker's signs of uncertainty and informational access.

Important future questions remain. If infants track accuracy (or consistency) within an individual, can they also track consensus (or inconsistency) across individuals? Does children's monitoring of the source differ across different domains of knowledge?

Might the sources of beliefs which are typically not universally held within a society be more memorable and more relevant? Are there certain beliefs for which the source is not tracked or is very difficult to remember? Finally, is there a psychological difference between beliefs that are based on consensus and beliefs that derive from a marked context or source? Among these and other issues that remain for future research, it will be important to discover what sorts of developmental precursors are found in infancy and early toddlerhood and the role that early socio-emotional experience plays in fostering selective trust.

References

Akhtar, N., Carpenter, M., & Tomasello, M. 1996. "The role of discourse novelty in early word learning." *Child Development,* 67: 635-645.

Baldwin, D. A. 1991. "Infants' contribution to the achievement of joint reference." *Child Development,* 62: 875–890.

Baldwin, D. A. 1993. "Early referential understanding: Infants' ability to recognize referential acts for what they are." *Developmental Psychology,* 29: 832–843.

Baldwin, D. A., Markman, E., Bill, B., Desjardins, N., Irwin, J., & Tidball, G. 1996. "Infants' reliance on a social criterion for establishing word–object relations." *Child Development,* 67: 3135–3153.

Baldwin, D. A., & Moses, L. 2001. "Links between social understanding and early word learning: Challenges to current accounts." *Social Development*, 10: 309-329.

Bartsch, K., & Wellman, H. M. 1995. *Children talk about the mind.* Oxford University Press: New York.

Birch, S., Luca, S., Frampton, K., Vauthier, S. & Bloom, P. 2005. "Children's assessments of what others know: the effects on learning." *Poster presented at the Biannual Society for Research in Child Development*, Atlanta, GA.

Bloom, P. 2000. *How Children Learn the Meanings of Words.* MIT Press: Cambridge.

Bretherton, I., & Beeghly, M. 1982. "Talking about internal states: The acquisition of an explicit theory of mind." *Developmental Psychology,* 18: 906–921.

Ceci, S. J., Ross, D. F., & Toglia, M. P. 1987. "Suggestibility of children's memory: Psycholegal implications." *Journal of Experimental Psychology: General,* 116 (1): 38-49.

Clark, E. V. 1993. *The lexicon in acquisition.* Cambridge: Cambridge University Press.

Clark, E. V. & Grossman, J. B. 1988. "Pragmatic directions and children's word learning." *Journal of Child Language,* 25: 1-18.

Clément, F., Koenig, M., & Harris, P.L. 2004. "The ontogenesis of trust." *Mind and Language*, 19: 360-379.

Corriveau, K.H. & Harris, P.L. 2007. "Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information." *Paper presented at the Biennial Meeting of the Society for Research in Child Development*, Boston, MA.

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.

Dawkins, R., & Krebs, J. R. 1978. "Animal signals: information or manipulation?" In J. R. Krebs and N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (pp. 282-309). Oxford: Blackwell Scientific Publications.

Fenson, L., Pethick, S., Renda, C., Cox, J., Dale, P., and Reznick, S. 2000. "Short-form versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics,* 21: 95-116.

Gelman, S. A., & Coley, J. D. 1990. "The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children." *Developmental Psychology,* 26: 796-804.

Gelman, S. A., Collman, P., & Maccoby, E. E. 1986. "Inferring properties from categories versus inferring categories from properties: The case of gender." *Child Development,* 57: 396-404.

Gelman, S. A., & Markman, E. M. 1986. "Categories and induction in young children." *Cognition,* 23: 183-208.

Gelman, S. A., & Markman, E. M. 1987. "Young children's inductions from natural kinds: The role of categories and appearances." *Child Development,* 58: 1532-1541.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. 1995. "Early object labels: The case for a developmental lexical principles framework." *Journal of Child Language,* 21: 125–155.

Gopnik, A., & Graf, P. 1988. "Knowing how you know: Young children's ability to identify and remember the sources of their beliefs." *Child Development,* 59: 1366-1371.

Graham, S. A., Stock, H. & Henderson, A.M.E. 2006. "Nineteen-month-olds' understanding of the conventionality of object labels versus desires." *Infancy, 9,* 341-350.

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

Henderson, A. M. E., & Graham, S. A. 2005. "Two-year-olds' appreciation of the shared nature of novel object labels." *Journal of Cognition and Development,* 6: 381-402.

Heyman, G. D., & Dweck, C. S. 1998. "Children's thinking about traits: Implications for judgments of the self and others." *Child Development,* 69**:** 392–403.

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.

Hirsh-Pasek, K., & Golinkoff, R. M. 1996. *The origins of grammar: Evidence from early language comprehension.* Cambridge, MA: MIT Press.

Hume, D. 1977. 'On Miracles' in *An Enquiry Concerning Human Understanding*. Cambridge, MA: Hackett Publishing Company.

Jaswal, V. K. 2004. "Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction." *Child Development,* 75: 1871-1885.

Jaswal, V. K., & Malone, L. S. in press. "Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility." *Journal of Cognition and Development.*

Jaswal, V. K & Markman, E. M. 2007. "Looks aren't everything: 24-month-olds'willingness to accept unexpected labels." *Journal of Cognition and Development,* 8: 93-111.

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.

Kalish, C. W. & Sabbagh, M. A. (Eds) 2007. *Conventionality in Cognitive Development: How Children Acquire Shared Representations in Language, Thought and Action.* Jossey-Bass: San Francisco, CA.

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: 179–203.

Koenig, M.A. & Harris, P.L. 2005. "Preschoolers mistrust ignorant and inaccurate speakers." *Child Development,* 76: 1261-1277.

Koenig, M.A., & Woodward, A.L. 2007. "The beginnings of mistrust: 24-month-olds' sensitivity to the prior accuracy of the source." *Manuscript under review.*

Lampinen, J. M., & Smith, V. L. 1995. "The incredible (and sometimes incredulous) child witness: child eyewitnesses' sensitivity to source credibility cues." *Journal of Applied Psychology,* 80: 621-627.

Lutz, D., & Keil, F. C. 2002. "Early understanding of the division of cognitive labor." *Child Development,* 73: 1073-1084.

Markman, E. M. 1989. *Categorization and naming in children.* MIT Press: Cambridge.

Markman, E. M., & Hutchinson, J. 1984. "Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations." *Cognitive Psychology,* 16: 1–27.

Mervis, C. B., Pani, J. R., & Pani, J. R. 2003. "Transaction of child cognitive-linguistic abilities and adult input in the acquisition of lexical categories at the basic and subordinate levels." In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 242–274). Oxford, England: Oxford University Press.

Miller, S. A. 2000. "Children's understanding of preexisting differences in knowledge and belief." *Developmental Review,* 20: 227 – 282.

Moore, C., Bryant, D., & Furrow, D. 1989. "Mental terms and the development of certainty." *Child Development,* 60: 167-171.

Moore, C., & Davidge, J. 1989. "The development of mental terms: Pragmatics or semantics?" *Journal of Child Language,* 16: 633-641.

O'Neill, D. K., Astington, J. W., & Flavell, J. H. 1992. "Young children's understanding of the role that sensory experiences play in knowledge acquisition." *Child Development,* 63: 474-490.

Onishi, K. H. & Baillargeon, R. 2005. "Do 15-month-old infants understand false belief?" *Science*, 308 (5719): 255-258.

Pasquini, E.S., Corriveau, K.H., Koenig, M.A., & Harris, P.L. in press. "Preschoolers monitor the relative accuracy of informants." *Developmental Psychology*

Pea, R. D. 1982. "Origins of verbal logic: spontaneous denials by two- and three-year olds." *Journal of Child Language,* 9: 597-626.

Reid, T. 1997. *An Inquiry into the Human Mind on the Principles of Common Sense*, Derek R. Brookes (ed.), Pennsylvania State University Press: University Park.

Rholes, W. S., & Ruble, D. N. 1984. "Children's understanding of dispositional characteristics of others." *Child Development,* 55: 550–560.

Robinson, E. J, Mitchell, P., & Nye, R. 1995. "Young children's treating of utterances as unreliable sources of knowledge." *Journal of Child Language,* 22: 663-685.

Robinson, E. J., & Whitcombe, E. C. 2003. "Children's suggestibility in relation to their understanding about sources of knowledge." *Child Development,* 74: 48-62.

Sabbagh, M. A., & Baldwin, D. A. 2001. "Learning words from knowledgeable versus ignorant speakers: Link between preschoolers' theory of mind and semantic Development." *Child Development,* 72(4): 1054-1070.

Sabbagh, M. A. 2007. "Judgments about knowledge enable young children to parse

signal from noise in social-information gathering." In S. Fitneva, S. Birch (Chairs), *Children's sensitivity to other's knowledge.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Boston, MA.

Sabbagh, M. A., Wdowiak, S. D. & Ottaway, J. M. 2003. "Do word learners ignore ignorant speakers?" *Journal of Child Language,* 30: 905-924.

Schafer, G., & Plunkett, K. 1998. "Rapid word learning by fifteen-month-olds under tightly controlled conditions." *Child Development,* 69: 309–320.

Sloutsky, V. M., & Fisher, A. V. 2004. "Induction and categorization in children: A similarity-based model." *Journal of Experimental Psychology: General,* 133: 166-188.

Taylor, M., Cartwright, B. S., & Bowden, T. 1991. "Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge?" *Child Development,* 62: 1334-1351.

Taylor, M., Esbensen, B. M., & Bennett, R. T. 1994. "Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned." *Child Development,* 65: 1581-1604.

Tomasello, M., & Barton, M. 1994. "Learning words in nonostensive contexts." *Developmental Psychology,* 30: 639–650.

Waxman, S. R., & Markow, D. B. 1995. "Words as invitations to form categories: Evidence from 12- to 13-month-old infants." *Cognitive Psychology*, 29: 257-302.

Welch-Ross, M. 1999. "Interviewer knowledge and preschoolers' reasoning about knowledge states moderates suggestibility." *Cognitive Development,* 14: 423-442.

Welder, A. N., & Graham, S. A. 2001. "The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties." *Child Development,* 72: 1653–1673.

Wellman, H., Cross, D., & Watson, J. 2001. "Meta-analysis of theory of mind development: the truth about false belief." *Child Development,* 72: 655 – 684.

Wimmer, H., Hogrefe, G., & Sodian, B. 1988. "A second stage in children's conception of mental life: Understanding informational accesses as origins of knowledge and belief." In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 93 – 108). New York: Cambridge University Press.

Woodward, A. L., Markman, E. M. & Fitzsimmons C. M. 1994. "Rapid word learning in 13- and 18-month-olds." *Developmental Psychology,* 30(4): 553-566.

Yuill, N. 1992. "Children's conception of personality traits." *Human Development,* 35: 265–279.

Bios:

Melissa Koenig is a developmental psychologist at the Institute of Child Development at the University of Minnesota. Her research focuses on cognitive development, early language development, social cognition and word learning. Her most recent work is aimed at children's understanding of conventional systems of knowledge.

Paul Harris is a developmental psychologist at the Graduate School of Education, Harvard University. He is interested in the development of cognition, imagination and emotion in early childhood. His most recent book is: *The Work of the Imagination* (Blackwell, 2000).