CHILD DEVELOPMENT

Child Development, November/December 2020, Volume 91, Number 6, Pages e1134–e1161

How Children Learn From Others: An Analysis of Selective Word Learning

David M. Sobel
Brown University

Zoe Finiasz
Cornell University

One way children are remarkable learners is that they learn from others. Critically, children are selective when assessing from whom to learn, particularly in the domain of word learning. We conducted an analysis of children's selective word learning, reviewing 63 papers on 6,525 participants. Children's ability to engage in selective word learning appeared to be present in the youngest samples surveyed. Their more metacognitive understanding that epistemic competence indicates reliability or that others are good sources of knowledge has more of a developmental trajectory. We also found that various methodological factors used to assess children influence performance. We conclude with a synthesis of theoretical accounts of how children learn from others.

A fundamental goal of developmental science is to describe what makes children such remarkably good learners. Research in cognitive development often focuses on how children might learn from the environment. For many years, there has been a great deal of interest in children's ability to learn from observations and actions (e.g., Gopnik & Wellman, 2012; Schulz, 2012; Sobel & Legare, 2014). Much of this work has been influenced by insights from social cognitive development, which suggest that children are not just constructing knowledge from internal processes as "stubborn autodidacts" (Harris, 2002, p. 341). Instead, the beliefs that children have about the world are acquired not just from observation and interaction with the environment, but from social transmission and communication (e.g., Csibra & Gergely, 2009; Harris, 2002, 2012; Vygotsky, 1978). The testimony of others provides key knowledge that cannot be obtained or would be difficult to obtain by observation alone (Harris & Koenig, 2006; Harris, Pasquini, Duke, Asscher, & Pons, 2006; see also Keil, 2010). Knowing from whom to learn, under what circumstances to learn, and then actually learning from others is the basis of cultural knowledge and transmission (e.g., Bergstrom, Moehlmann, & Boyer, 2006; Harris & Koenig, 2006; Kline, 2015; Mascaro & Sperber, 2009).

This research was supported by NSF (1661068 to David M. Sobel). We thank D.J. Williams for help with coding, Kathleen Corriveau, Paul Harris, Vikram Jaswal and Melissa Koenig for help with our literature review.

Correspondence concerning this article should be addressed to David M. Sobel, CLPS Department, Brown University, 190 Thayer St., Providence, RI 02912. Electronic mail may be sent to dave_sobel@brown.edu.

"Learning from others" has been conceptulized in a number of ways. There is a long literature in cognitive development on the importance of explanations for learning (Brewer, Chinn, & Samarapungavan, 1998; Keil & Wilson, 2000), so much so that some have argued that children have a "drive for explanation" (Gopnik, 1998, p. 109), which motivates what and how they learn. How children generate explanations is often conceptualized as a measure of their knowledge (e.g., Schult & Wellman, 1997; Shultz, Wells, & Sarda, 1980; Sobel, 2004). Children's learning also benefits from generating their own explanations (e.g., Chi, De Leeuw, Chiu, & LaVancher, 1994; Crowley & Siegler, 1999; DeLeeuw & Chi, 2003; Legare & Lombrozo, 2014; Lombrozo, 2006) and from the interaction between explaining information and exploring the environment (see Legare, Sobel, & Callanan, 2017, for a review). Moreover, how children learn from others often relates to the dichotomies between instruction and discovery (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). To what extent is children's learning influenced by the interaction between direct instruction—explanations and information generated by others—and guided interaction in which others scaffold children's actions (Klahr, 2000; Klahr & Nigam, 2004; Kuhn, 1989; Kuhn & Dean, 2005; Weisberg, Hirsh-Pasek, Golinkoff, Kittredge, & Klahr, 2016)? Direct instruction, and even overhearing information (Akhtar, 2005), has clear importance for learning, even if there are tradeoffs

 0009-3920/2020/9106-0027
DOI: 10.1111/cdev.13415

in the extent to which children are engaged by the learning process (Bonawitz et al., 2011).

This brings up a critical question, which is the topic of this review: Do children simply believe what they are told by others or do they learn from others in a more judicious manner? That is, are children wholly credulous of others' information or more selective in their learning? Young children have traditionally been thought to be overly credulous; Piaget (1930), for example, suggested that it was not until relatively late in development that children could be skeptical of others' testimony. Piaget's assumption, however, has not been supported; it is now widely accepted that children are not just credulous of others' testimony but rather are selective regarding from whom they learn. Children can track the reliability of other informants and use that information to make judgments about their epistemic competence and whether to learn from them (Clément, Koenig, & Harris, 2004; Koenig, Clément, & Harris, 2004; Koenig & Harris, 2005). Selective learning is now seen as an important part of social and cultural learning, and there has been over 15 years of research examining the extent to which children (mostly 2- to 6-year-olds, the age group on which we will focus) track others' reliability, make inferences about others' epistemic competence, and use their judgments of epistemic competence to learn from others.

Given that children can make selective inferences, an open question concerns the kinds of inferences they make and how they make them. Judgments about epistemic competence could be limited to deciding whether to use the information provided or could extend to dispositional traits about the informant (i.e., not only will I not use this person's information, but this person is a bad source of knowledge and should not be relied on the future). Judgments about epistemic competence could also involve predictions about who has particular pieces of knowledge and thus who should be relied on, or could be limited to simply using information without rationales for why one used that knowledge.

The extent to which children generalize inferences from others' epistemic competence bears on predictions made by different theoretical accounts of selective learning. One way of interpreting Piaget's (1930) arguments for credulity is that young children have trouble differentiating between the physical and mental worlds, and thus potentially struggle to appreciate differences between what others say and what they should believe (see Woolley & Ghossainy, 2013, for a review). This credulity is rooted in various philosophical accounts, most notably Reid (1764/2000), who suggested that trust was a natural part of social interaction and that evidence gained from testimony was no different to evidence gained by direct observation. Such trust is reflected in psychological theories such as Jaswal, Croft, Setia, and Cole (2010), who argue that children's "willingness to believe what they are told stems from a specific, highly robust bias to trust testimony." (p. 1546). Children initially believe others' information, and become selective in their learning. The onset of selective learning might depend on general cognitive capacities, like inhibitory control (Jaswal et al., 2014) or their general metacognitive understanding of skepticism (Mills, 2013). But both these accounts suggest that early on, children show limited selective learning capacities and there is development, particularly during the preschool years. This robust bias to trust also serves as part of the origin of theories of selective learning based on more interpersonal trust (e.g., Harris, Koenig, Corriveau, & Jaswal, 2018).

An alternate psychological account of selective learning potentially has its origins in philosophical theory articulated by Coady (1992), who suggested that information generated by others is only justified to be treated as true if it is supported by other beliefs acquired by first-person means. Children might have a more rational basis for trusting others, given that they treat information generated by others as a form of data that is verifiable (Sobel & Kushnir, 2013). What is critical for development on this account is children's own knowledge—knowledge of the world affords both the ability to verify others' information as well as an understanding of when to generalize accuracy broadly or narrowly. This view suggests that there is little development of children's selective learning capacities and much more development of specific knowledge that might affect selective learning in individual domains.

While these different reviews of the literature offer different perspectives, there are few integrations of these findings through meta-analysis of the current literature. This is the goal of this manuscript. Analyses of this form have been informative in the theory of mind (ToM) literature for synthesizing findings, particularly within the false belief literature, by both describing developmental trajectories as well as what manipulations have been meaningful to children's performance (Wellman, Cross, & Watson, 2001). Given the large number of published

studies on selective word learning, we attempted to take the same approach.

Our research questions center on how children engage in selective learning. First, what inferences can children make about the reliability of others as potential sources of knowledge? Second, do those inferences have distinct developmental trajectories? Finally, what cues do children use to make inferences about others' reliability as potential sources of knowledge—that is, to what extent is children's selective learning influenced by how information is presented as opposed to the nature of the information children must use to reason about others' reliability?

There are many domains of knowledge that require learning from others. How and whether we imitate, how we develop religious beliefs or beliefs about unobservable entities, and how we acquire scientific knowledge all require (or at least are deeply aided by) information communicated to us in the form of explanation, instruction, or guidance. Because of the wide range of methods and ways in which selective learning has been investigated, we needed to restrict our analysis in certain ways. We chose selective word learning because it is paradigmatic of the selective learning literature. There is a conventional, but arbitrary relation between the phonological structure and the lexical semantics of one's language (e.g., Bloom, 2000; de Saussure, 1966). Word learning (save perhaps onamonapia), thus, cannot exist in a vacuum; it must rely on others' information. Becoming a member of a linguistic community cannot be done in the absence of said community. Moreover, the ability to infer the referent of a novel label to a novel object has been the most predominant method used to investigate selective learning. But we also restricted our analysis to selective word learning for two practical reasons. First, many studies on children's selective word learning use common methods, making the comparison among the studies more cohesive. Second, as we will describe in the following section, there are many cues to trusting others, which have all been potentially applied to selective word learning in particular. We can use the literature on trust to guide whether particular cues affect children's selective word learning.

Types of Questions Used in Studies of Selective Word Learning

In one of the first measures of selective word learning, Koenig et al. (2004) examined how children use others' history of accurately generating appropriate labels of familiar objects to endorse novel labels of novel objects. They introduced 3- and 4-year-olds to two novel informants. One labeled familiar objects accurately. The other labeled the same objects inaccurately. They asked *Explicit Judgment* questions about these informants—whether one informant was either a good or bad labeler of the familiar objects. They also asked *Endorse* questions: They presented children and the informants with novel objects and asked the informants to label them. The informants each generated a different novel label (e.g., one called the novel object a dax, the other a wug). Children were then asked to label the novel object dax or wug. They found that children were relatively good at stating that the previously accurate informant was good at labeling things (and that the previously inaccurate informant was bad at labeling things). They also found that children endorsed the accurate informant's labels more often than the inaccurate informant's labels.

Koenig and Harris (2005) expanded on this paradigm by introducing a third dependent measure, in which children were asked a more metacognitive question about which informant possessed particular kinds of knowledge. After establishing that one informant was an accurate labeler and another was an inaccurate labeler, they asked *Ask* questions to preschoolers (in addition to Endorse and Explicit Judgment questions): whom did the children want to ask for a novel label. Children were more likely to point to the accurate informant as someone they would want to ask. In this experiment, there was some development: Four-year-olds were better at answering Ask and Endorse questions than younger children. Koenig and Harris also pointed out another important facet of selective learning, which is that there is the potential for the generalization of knowledge. In a subsequent experiment, they showed that children would use informants' accuracy at labeling objects to make inferences about information provided by those informants about novel objects' functions.

All the papers we reviewed on selective word learning implemented a subset of the Ask, Endorse, and Explicit Judgment questions. How children respond to these questions will be the focus of our analyses. As such, it is important to consider what each question type specifically asks of children, and what inferences each question type requires children to make.

Endorse Questions

Endorse questions—typically of the form "What is this <object> called?" require children to

generate a label for a novel object or approve a label generated by an informant. These questions generally get at the semantic inference that children are being asked to make during word learning: What is an acceptable label for an object, which would be shared by members of my linguistic community? In some cases, endorse questions could also involve asking children what label they would use in the future if they needed to refer to that object.

Explicit Judgment Questions

Explicit judgment questions—typically of the form "Is this person a good or bad (or silly) labeler?" or "Who is better at naming things?"—have often been considered control questions. The expectation was that children might respond correctly on these questions because they recognized that the informants were indeed accurate or inaccurate during the familiarization phase of the procedure (during which informants generated information children would use to infer their subsequent veracity). The question was whether they would use that accuracy to make subsequent inferences (in the form of Ask or Endorse questions).

Treating the explicit judgment questions as a control question raises some points for discussion. The first is that several researchers have used correct responses on explicit judgment questions as an inclusion criterion, excluding children from analyses if they answer one or more of these questions incorrectly. On this view, the explicit judgment question is considered a memory question for the information generated by the informants during the familiarization phase. Incorrect responding indicates a failure to remember the information the informants generated or a failure to attend to their information, either of which would make responses to any other question unrepresentative.

Such an approach is consistent with that used by Perner Leekam and Wimmer (1987) on the unexpected contents false belief task. Their control question was to ask what was really in the deceptive container. Children had to respond correctly to in order to make a mental representation about their own or others' mental states. Performance on this control question influences how children respond to the test questions about others' false beliefs (Sobel & Austerweil, 2016). Is such an approach warranted here? Does the explicit judgment question act as a memory control?

The second is whether children find it pragmatically odd to have to respond to this question, and then other questions that could directly contradict the answer to this question. If the explicit judgment question is asked first, then does it affect how children respond to subsequent questions? That is, if children respond that one informant is a better labeler than the other, and then children are asked to endorse one of the informants' novel labels, do they chose the same informant because they genuinely believe that this information is correct or because they have just been asked who is a better labeler, and to not pick that informant's labels would register a contradiction? Similarly, if the explicit judgment question is asked last, does the presence of the other questions asking about the informants influence performance on the explicit judgment question? We want to examine both these concerns in our analysis.

But, there is another potential interpretation of the explicit judgment question, beyond as just a control for the other test questions. In order to make inferences about the labels for objects, it is important to show that children appreciate that there is a standard to which the informant's information is compared, or that children can make a trait generalization about the sources of information (i.e., suggesting that a shortcut for thinking about informants who generate inaccurate labels is that they are just bad labelers). On this view, the explicit judgment question considers a unique aspect of children's selective learning—the trait inference they make based on the informants' information. This could reveal whether they would use that information in the future, beyond just the endorse question, which involves whether they use that information in this instance.

Ask Questions

The ask question requires children to state whom they think knows the label for the novel object. These questions are often asked in conjunction with the endorse question (although always before it, for practical reasons). As such, the same concern about the explicit judgment question applies to the ask question—whether performance on it or on other questions is affected by whether it is asked.

But beyond this concern, another question is whether the ask question is interchangeable with other questions, and particularly the endorse question. Many researchers combine the results of the ask and endorse questions in their analyses. But while the endorse question seems to ask about the

specific label for the object, the ask question seems more of a metacognitive question about children's understanding of the knowledge states of the informants. An open question is whether ask questions are more difficult or have a different developmental trajectory.

Factors to Making Selective Inferences About Word Meaning

Studies of selective word learning have investigated several different questions. The first is whether children's selective word learning capacities develop. This question has generated much interest and debate. Different investigations of selective word learning posit different answers to this question. The early studies by Koenig, Harris, and colleagues have suggested developmental differences between 3- and 4-year-olds. Jaswal and colleagues (Jaswal, 2010; Jaswal et al., 2010) have suggested that very young children might be highly credulous in what they believe when hearing testimony from others (see also Mills, 2013) and that children develop the capacity to be more skeptical in their word learning. Such a hypothesis was supported by findings by Krogh-Jespersen and Echols (2012) who showed that 2-year-olds learned novel labels equivalently from informants who were accurate, inaccurate and ignorant about the labels for familiar objects. In contrast, several studies (e.g., Brooker & Poulin-Dubois, 2013; Koenig & Woodward, 2010; Luchkina, Sobel, & Morgan, 2018) demonstrated that by age 2, children were selective in their word learning. While some have pointed to methodological differences that might account for these differences (e.g., Vanderbilt, Heyman, & Liu, 2014), it remains unclear how these different methods potentially affect learning from others selectively. Moreover, it remains unclear whether these differences affect the three different questions under discussion (ask, endorse, and explicit judgment) differently.

The development of selective word learning might also be influenced by the way in which information is presented. Some researchers present live informants to children, whereas others use video presentations, photos of informants, or puppets. Performance on these selective learning measures might differ when children are faced with having to reject information generated by a live informant as opposed to a representation of that informant such as a video or photograph (see Deloache, 2000, for a review of children's appreciation of such psychological distancing effects). These methodological differences all have the potential to affect children's responses on these measures.

Beyond methods used to measure selective learning that might affect children's performance, there is also a question about what cues influence children's selective learning. At this point, we want to distinguish selective learning from selective *trust*, and restrict our analysis to the former. Trust is a complex psychological concept that requires both epistemic and emotional inferences and relies on a host of cognitive and social developmental factors (e.g., Bernath & Feshbach, 1995; Rotter, 1967, 1971). Trust surely influences learning. But the evaluation and decision to rely on others' information can potentially be made independently of the decision to trust. That is, I might not trust certain individuals, but can still deem them epistemically competent and be willing to learn from them in some circumstances. Some reviews distinguish *epistemic trust* from *interpersonal trust* to make the distinction we are making here (e.g., Harris et al., 2018; Sobel & Kushnir, 2013). Epistemic trust refers to the extent to which one acquires the epistemic knowledge communicated to them. Interpersonal trust refers to the extent to which one likes another or feels associated with them, such that an emotional reaction is evoked.

This distinction is important because certain cues could potentially affect judgments of both learning and trust, whereas other cues that have been investigated in the selective word learning literature come from the literature on trust, but have no bearing on judgments about the epistemic competence of an informant. That is, there are some cues that affect children's judgments about others' epistemic competence, whereas other cues might be used to make judgments about trust—the interpersonal relationship children might have with another person—but not necessarily bear on that individual's epistemic knowledge. For example, adults judge that attractive faces are more trustworthy (Todorov, Pakrashi, & Oosterhof, 2009). Attractiveness, however, does not indicate epistemic competence. Although there is not a rational reason to use this cue as a basis for learning, there are some findings that posit children are sensitive to this cue (Bascandziev & Harris, 2014).

One way of considering this difference is to acknowledge that there are numerous characteristics that can be assessed about an individual prior to their generating any information about the knowledge they possess. An important question is whether children's selective learning is more influenced by cues that specifically indicate the epistemic competence of an informant or cues that are merely associated with such cues (or with positive

valence generally construed). This distinction, however, is not simple. Jaswal and Neely (2006) showed that when given a contrast between an adult and child informant, preschoolers will rely on the adult as a source of information for novel labels for novel objects in the absence of any other cues to these informants' reliability. Children might trust that adults are better sources of knowledge than other children. But children can learn that this is not the case by establishing the adult informant to be inaccurate and the child informant to be accurate at generating labels for familiar objects. We considered how children reasoned about cues that unambiguously indicated epistemic competence (accuracy or expertise) as opposed to cues that could be discounted given the presence of more unambiguous cues. Of course, in many studies, these cues are not mutually exclusive, but rather tested together—a point of consideration that we make in this analysis.

Relations to Other Meta-Analyses

The present investigation has three goals. First, we want to investigate the developmental trajectory of selective word learning, specifically focused on whether there are differences among the three types of questions that have been commonly used in this literature. This will potentially document similarities or differences among these questions, and what they each mean for children's selective learning more generally. Second, we want to consider how various methodological factors that have been used throughout the literature affect performance. Our focus here is on documenting how who the informants are and what the pragmatic cues they generate affect children's performance. Finally, we examine what cues children might use to make inferences about selective learning, and how those cues differ in relative strength, both on their own and when they are consistent with one another or in conflict.

Tong, Wang, and Danovitch (2019) published a set of meta-analyses on the paradigm originally described by Koenig et al. (2004). In their first analysis, they showed that between the ages of 3–6, children were more likely to ask and endorse individuals who demonstrated a greater degree of epistemic reliability, with age 4 appearing to be a key transition point. In their second analysis, they showed that children would use social characteristics—such as cues to social trust documented above, like consensus, attractiveness, or in-group membership—as the basis of responding to ask and endorse questions, but there were no clear age effects. Finally, in their third analysis, they examined cases where these social characteristics conflicted with cues to epistemic knowledge. They found that children did not show an overall preference on the ask questions, but were more likely to endorse epistemic cues over social characteristics. There was a weak effect of age in this analysis, such that older children were marginally more likely to do this than younger children.

The Tong et al. (2019) meta-analyses offer key conclusions regarding the literature on selective word learning. The analysis we present here uses a different meta-analytic technique and engages in different set of coding. In this way, we can both expand on questions they did not consider in their analysis, as well as complement the work they present to see whether similar findings emerge across techniques. There are several key points to highlight. Several are methodological. First, Tong et al. only considered cases in which children learned from two informants, and did not contrast learning from a single informant. Procedures that only present a single informant tend to be more ecologically valid, as it is rare to see two informants in the real world generate different labels for the same object in quick succession. Including these analyses (as well as whether they yield different results) expands how researchers might interpret the literature on selective word learning. Second, the present analysis focuses on the ask and endorse questions as well as the explicit judgment question, and whether including this question matters as a control. Including this question allows us to consider whether the way researchers have controlled experiments on selective word learning affects the results. Third, Tong et al. (2019) treated age categorically (3-year-olds, 4-year-olds, and 5- to 6-year-olds) in many of their analyses. Using a more continuous approach to age might result in different findings.

But we also want to highlight a theoretical concern in this literature, which Tong et al. rightly point out in their discussion: "the line between epistemic and social characteristics can be blurry" (p. 7). There are cases where competing social characteristics have no obvious bearing on informants' epistemic competence, such as in cases of informants differing in attractiveness. In other cases, differences in social characteristics might indicate differences in epistemic knowledge. For example, certain in-group/out-group distinctions, such as speakers who provide similar lexical information but in native versus non-native accents, might signal both in-group membership as a social category

as well as membership in a linguistic community that signals epistemic information (native speakers might know more words than non-native or second-language learners). We attempt to consider this in two ways. First, we look at cases where epistemic and nonepistemic cues are in conflict (similar to Tong et al.), but also where they are consistent with one another to see if that strengthens the use of a particular informant's information. Second, where we can, we try to examine individual nonepistemic cues to see whether they differ in their effect on children's judgments.

Finally, an important difference between the approach taken here and that of Tong et al. (2019) concerns how the paradigms for testing selective word learning are interpreted. We have specifically used the term "selective word learning" in our review, instead of the more general term "selective trust." Theoretically, we would like to draw a distinction between selective learning and trust, a point we will return to in the discussion, by examining how these analyses shed light on mechanisms for selective word learning.

Analysis

The analysis reported here was inspired by the analysis conducted by Wellman et al. (2001) and others (e.g., Sobel & Austerweil, 2016; Wellman & Liu, 2004) on children's theory of mind. We reviewed a set of papers on children's selective word learning, calculating the proportion of children who relied on different sources of knowledge to respond to test questions. Like Wellman et al. (2001), we used age group and condition within an experiment, instead of participants or experiments themselves as the unit of analysis (following Glass, McGaw, & Smith, 1981). But, because children in selective word learning procedures tend to answer different kinds of questions, we also considered question type as a basic unit of analysis.

Literature Review

We began our investigation by amassing a set of papers on selective word learning. We ran this literature search in December 2017 and included published work if it was available at this time (e.g., through first-view mechanisms online). We initially used the Google Scholar citation function to investigate what papers cited Koenig et al. (2004) and Koenig and Harris (2005) to construct our initial list. We also consulted the reference lists of two well-known reviews (Mills, 2013; Sobel & Kushnir, 2013) to add to this list. We supplemented this list by asking a set of experts (Kathleen Corriveau, Paul Harris, Melissa Koenig, and Vikram Jaswal) to add any additional paper they thought was relevant and by sending a request for findings on the *cogdev-soc* listserv (in December, 2017) for manuscripts that would be appropriate.

We initially read the papers to ensure that we could extract meaningful information from the methods and results sections and that the papers met the inclusion criteria of our study. We considered several criteria for inclusion during this initial reading. First, although selective learning is appropriate in many other domains (e.g., imitation, factual knowledge, invisible entities, religious thinking), we restricted our analysis to studies that used word learning as a dependent measure in at least one condition of an experiment. In the general discussion we will consider the generalizability of this analysis to other domains of selective learning.

Second, we did not consider any study that failed to report data sufficient to be included in the analysis. For example, to be included in our analysis, the study needed to report children's age (both mean and a measure of variance) and performance on individual test questions that fit into at least one of the Ask, Endorse, and Explicit Judgment structure. In some cases, we sought additional information from authors, but this information was not always possible to obtain, and when authors did not respond, we did not include these data in the analysis.

Third, we excluded any condition in which more than 40% of the total sample was excluded, which was consistent with Wellman et al.'s (2001) analysis. Only one study within one article we considered met this criterion.

In sum, our analysis included 63 published articles that totaled 105 separate studies and 666 groups/conditions within those studies. This represents the results of 6,525 individual participants. A list of the studies is shown in Table 1. The majority of the studies we investigated involved children between the ages of 2–5. This is not to say that by the age of 5, children know all there is to know about selective word learning. Indeed, there were many cases in which 5-year-olds performance was not at ceiling levels. Our goal is not to suggest that at any one particular age children can engage in particular kinds of selective word learning. Rather, our goal is to show that there are differences in the ways that selective word learning is measured and that those differences might indicate what and how children are learning.

Table 1
List of Studies included in Analysis

Authors and year	Experiments in paper	Groups/conditions by question included	Categorization of groups/conditions by question included
Bascandziev and Harris (2014)	1	4	4 Nonepistemic: perceptual valence
Bascandziev and Harris (2016)	1	12	4 Epistemic 4 Nonepistemic: perceptual valence 4 Both in conflict: perceptual valence
Bernard, Proust, and Clément (2015)	2	9	6 Both in conflict: consensus 3 Both no conflict: consensus
Birch and Bloom (2002)	2	12	12 Nonepistemic: linguistic
Birch, Vauthier, and Bloom (2007)	2	4	4 Epistemic
Brosseau-Liard, Cassels, and Birch (2014)	1	4	2 Nonepistemic: confidence 2 Both in conflict: confidence
Brosseau-Liard et al. (2015)	1	2	1 Epistemic 1 Nonepistemic: perceptual valence
Canfield, Saudino, and Ganea (2015)	1	4	4 Epistemic
Chen, Harris, and Corriveau (2012)	2	18	18 Nonepistemic: consensus
Clément, Bernard, Grandjean, and Sander (2013)	1	9	9 Nonepistemic: type of informant
Corriveau, Fusaro, and Harris (2009)	2	12	12 Nonepistemic: consensus
Corriveau and Harris (2009a)	2	30	12 Nonepistemic: type of informant 9 Both in conflict: type of informant 9 Both no conflict: type of informant
Corriveau and Harris (2009b)	2	38	38 Epistemic
Corriveau, Kinzler, and Harris (2013)	2	15	2 Epistemic 3 Nonepistemic: linguistic 5 Both in conflict: linguistic 5 Both no conflict: linguistic
Corriveau and Kurkul (2014)	2	8	8 Epistemic
Corriveau, Kurkul, and Arunachalam (2016)	2	4	4 Epistemic
Corriveau et al. (2009)	1	18	18 Epistemic
Corriveau, Pickard, and Harris (2011)	3	16	16 Epistemic
Danovitch and Alzahabi (2013)	3	18	18 Epistemic
Doebel and Koenig (2013)	1	4	4 Nonepistemic: type of informant
Doebel, Koenig, and Rowell (2016)	2	21	21 Epistemic
Einav and Robinson (2010)	3	13	13 Epistemic
Einav and Robinson (2011)	2	4	4 Epistemic
Elashi and Mills (2014)	2	32	16 Both in conflict: type of informant 16 Both no conflict: type of informant
Fusaro, Corriveau, and Harris (2011)	1	2	1 Epistemic 1 Nonepistemic: perceptual valence
Fusaro and Harris (2008)	1	6	6 Epistemic
Guerrero, Elenbaas, Enesco, and Killen (2017)	1	1	1 Nonepistemic: consensus
Hermes et al. (2016)	1	4	2 Epistemic 2 Nonepistemic: perceptual valence
Jaffer and Ma (2015)	1	2	2 Nonepistemic: perceptual valence
Jaswal and Neely (2006)	1	4	2 Nonepistemic: type of informant 1 Both in conflict: type of informant 1 Both no conflict: type of informant
Kim and Harris (2014a)	1	2	2 Nonepistemic: other
Kim and Harris (2014b)	1	2	2 Nonepistemic: other
Kim, Paulus, and Kalish (2016)	1	10	10 Epistemic
Koenig et al. (2004)	1	6	6 Epistemic
Koenig and Harris (2005)	3	18	18 Epistemic
Koenig and Jaswal (2011)	2	12	12 Epistemic
Koenig and Woodward (2010)	4	11	11 Epistemic

Table 1
Continued

Authors and year	Experiments in paper	Groups/conditions by question included	Categorization of groups/conditions by question included
Koenig and Woodward (2012)	1	8	8 Epistemic
Kondrad and Jaswal (2012)	1	8	8 Epistemic
Krogh-Jespersen and Echols (2012)	1	10	10 Epistemic
Kushnir and Koenig (2017)	2	5	5 Epistemic
Landrum, Mills, and Johnston (2013)	3	16	6 Epistemic 8 Both in conflict: type of informant 2 Both no conflict: type of informant
Lane, Wellman, and Gelman (2013)	1	18	18 Nonepistemic: type of informant
Lucas, Lewis, Pala, Wong, and Berridge (2013)	1	15	15 Epistemic
Luchkina et al. (2018)	2	4	4 Epistemic
Luu, de Rosnay, and Harris (2013)	1	27	27 Epistemic
MacDonald, Schug, Chase, and Barth (2013)	3	9	3 Epistemic 3 Both in conflict: type of informant 3 Both no conflict: type of informant
McDonald and Ma (2015)	2	2	2 Nonepistemic: perceptual valence
Nurmsoo and Robinson (2009)	3	8	2 Epistemic 4 Nonepistemic: other 2 Both no conflict: other
Pasquini et al. (2007)	2	30	30 Epistemic
Rakoczy, Hamann, Warneken, and Tomasello (2010)	1	2	2 Nonepistemic: type of informant
Rakoczy, Warneken, and Tomasello (2009)	1	12	12 Epistemic
Robinson, Fox, and Einav (2013)	2	6	6 Nonepistemic: linguistic
Sabbagh, Wdowiak, and Ottaway (2003)	1	3	3 Epistemic
Scofield and Behrend (2008)	2	4	4 Epistemic
Sobel and Corriveau (2010)	1	10	10 Epistemic
Sobel and Macris (2013)	2	10	10 Epistemic
Sobel, Sedivy, Buchanan, and Hennessy (2012)	1	12	12 Epistemic
Stephens and Koenig (2015)	1	6	6 Epistemic
Taylor (2013)	1	12	6 Nonepistemic: type of informant 3 Both in conflict: type of informant 3 Both no conflict: type of informant
Vanderbilt et al. (2014)	4	12	12 Epistemic
Vanderbilt, Ochoa, and Heilbrun (2018)	3	20	20 Epistemic
Yow and Li (2018)	1	6	6 Nonepistemic: linguistic

Coding

Each age group/condition was coded for the proportion of appropriate response to the particular test questions asked (i.e., ask, endorse, or explicit judgment questions). If children were asked more than one of these questions, their responses were entered as multiple rows in the analysis. *Appropriate* was defined as appealing to the informant who was more epistemically competent, who had more positive valence (if epistemic competence was not possible to determine), or who was more in line with the hypotheses of the particular study (if neither of those factors were possible to determine). If more than one of these test questions were asked, then these were considered separately.

Following Wellman et al. (2001), proportion correct data were analyzed via a logit transformation to determine odds ratios more naturally and to eliminate concerns with analyzing proportional data. In the rare cases where a particular group/condition responded at 100% correct, we entered this value to the logit transformation as 0.99 instead of 1 (as a value of 1 would produce an invalid transformed value). The reported analyses do not change if these conditions are simply excluded from the analysis.

In addition to proportion correct as our dependent measure, we coded the following information

for each condition, which serve as variables in our analyses:

1. Year of paper publication: The year in which a paper was published should have no bearing on performance.
2. Number of participants in the group/condition: Measuring the sample size of each group serves two purposes. First, the specific number of participants in each group or condition allows us to investigate the extent to which researchers have engaged in researcher degrees of freedom regarding sample size choice. The absence of a relation between sample size and performance would suggest stability in the findings.
3. Mean age of the participants in the condition.
4. Informant type: This was a categorical code reflecting who the informant was. The categories included: (a) Photographs of informants; (b) Videos of informants; (c) Puppets; (d) Live actors. In some cases, researchers used combinations of methods (i.e., they saw puppets during familiarization, but photographs of the puppets at test). These represented a small number of cases, and we did not include them in the analysis of this variable.
5. Number of informant groups during familiarization: One or Two. In most cases, an informant group was an individual informant; in some cases, particularly when researchers were studying conformity, informant groups consisted of multiple participants who all said the same thing, or who responded together as a group. This measure determined if children saw one or two sources of information during familiarization.
6. Whether informants conflicted in labeling the same object during familiarization: This categorical variable reflected whether participants heard multiple informants generate different information about the same object during familiarization (i.e., one informant call a shoe a "shoe" whereas the other called the same shoe a "duck"). This code was only appropriate when there was more than one informant group present as coded in #5.
7. Whether an explicit judgment question was asked: This was a categorical variable. If an explicit judgment question was asked, we also considered whether it was asked before or after the other test questions (or both if it was asked twice). We also considered whether success on the explicit judgment question was used as an inclusion criterion for analysis of the other test questions.
8. The type of knowledge presented during familiarization: This categorical variable reflected whether the information presented to children during familiarization specifically indicated the epistemic competence of the informants or not. Experiments that varied the accuracy or expertise of the informants were considered indicative of epistemic competence. All other variables were considered nonepistemic manipulations. A nonepistemic manipulation was indicated by whether coders judged that it would be overruled by the presence of a particular epistemic one—accuracy. That is, coders judged whether they thought children should not use this cue if it was in contrast with the accuracy of the informant.

We further subdivided nonepistemic manipulations based on the type of manipulation performed by the investigators. These subcategories included (a) *Linguistic*—manipulations in which speakers used different language patterns. This included contrasting native versus non-native accents, correct versus incorrect syntax, or appropriate versus inappropriate pragmatic constructions. (b) *Perceptual Appearance*—manipulations that varied the visible salience of the informants, such as making one of them more attractive than the other. (c) *Consensus*—manipulations that varied the number of informants generating distinct types of information, such as contrasting information from a group of informants as opposed to a single informant. (d) *Confidence*—manipulations that varied the confidence with which informants generated information. (e) *Type of Informant*—manipulations in which informants were members of contrasting groups (adult vs. child, in-group vs. out-group, nice vs. mean). (f) *Other*—other manipulations that did not necessarily indicate the epistemic competence of the informants. These subgroups are indicated in Table 1.

We also considered whether the group or condition only manipulated an epistemic or nonepistemic factor or both. For those that manipulated both, the two factors could be consistent with one another or could be in conflict. For example, when a nonepistemic factor like speaking in a native accent was paired with an epistemic factor like accuracy, whereas a speaker with a non-native accent was paired with inaccuracy, we considered these factors consistent. A condition in which a non-native accented speaker was accurate and the native accented speaker was inaccurate, would be considered in conflict.

All these codes were initially coded by the second author. A subset of these papers (10 of them,

representing 120 questions/conditions or 18.02% of the total sample) was also coded independently by an undergraduate research assistant, blind to the purpose of the review. Agreement was 96%. Disagreements were resolved through discussion between the two authors.

Results

For all analyses, we considered our transformed proportion correct score as the dependent variable, and report effect sizes as Odds Ratios where appropriate (and as partial eta-squared when not). Our initial analysis concerned the role of age in selective word learning among the three types of questions that we analyzed (Ask, Endorse, Explicit Judgment). The effect of age was a significant predictor of performance, $R^2 = .025$, $F(1, 664) = 17.00$, $p < .001$, OR = 1.17. This analysis, however, combines results of the three questions together. Performance among the questions differed, $R^2 = .087$, $F(2, 663) = 31.57$, $p < .001$, $\eta_p^2 = .086$, and there was a significant interaction between age and question type, $R^2 = .085$, $F(2, 663) = 30.86$, $p < .001$, $\eta_p^2 = .085$. Figures 1a–1c shows the patterns of data for the Ask, Endorse, and Explicit Judgment questions, respectively.

Overall, children performed better on the Explicit Judgment question than on the Ask or Endorse questions, $B = -0.54$ and -0.65, $SE = 0.10$ and 0.09, Wald $\chi^2(1) = 26.05$ and 55.19, both $p < .001$, but did not differ between the Ask and Endorse questions, $B = -0.12$, $SE = 0.09$, $\chi^2(1) = 1.57$, $p = .21$. Moreover, not all the question types significantly correlated with age. The mean age of participants correlated with performance on the Ask question, $R^2 = .081$, $F(1, 144) = 12.74$, $p < .001$, OR = 1.32 and the Explicit Judgment question, $R^2 = .028$, $F(1, 172) = 4.88$, $p = .028$, OR = 1.18, but not on the Endorse question, $R^2 = .004$, $F(1, 344) = 1.22$, $p = .27$. This finding is consistent with the significant interaction between age and question type.

Inspection of Figures 1a–1c also reveals that the intercepts of the three regression lines are different. The intercepts for the Endorse and Explicit Judgment are similar (0.48 and 0.41 respectively), whereas the intercept for the Ask question is lower (−0.44). This is consistent with the main effect of question type and interaction between question type and age reported earlier. Intuitively, these results suggest that performance on the Endorse and Explicit Judgment questions start at similar levels, but performance on the Explicit Judgment questions improves with age, whereas performance

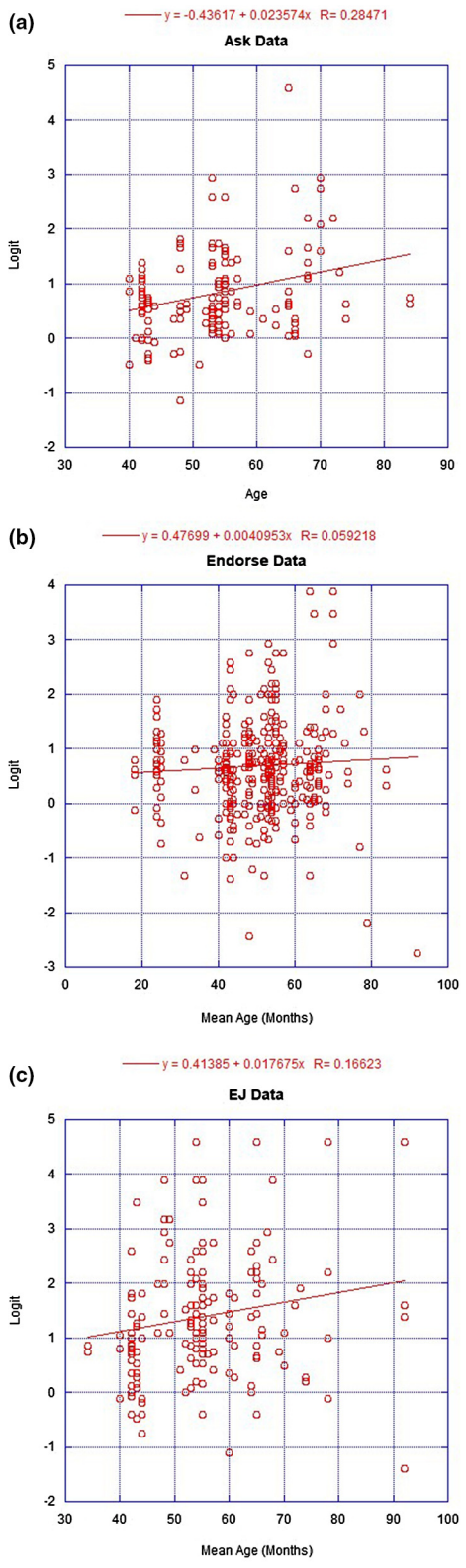

Figure 1. Relation between age and proportion correct as represented by the logit function used in the analysis for the (a) ask question, (b) endorse question, and (c) explicit judgment question.

on the Endorse question is stable between the ages these papers investigated (mostly 2- to 6-year-olds). Performance on the Ask question starts at a lower level than either of the other two questions early in development, but improves with age, similar to the Explicit Judgment question.

Given the differences in responses among the three questions, the remaining analyses were conducted on these questions separately to reveal whether there were significant effects on the variables we examined. Tables 2–4 list the results of our analysis for the explicit judgment, ask, and endorse questions, respectively. We will examine findings of methodological and theoretical significance, highlighting both significant and nonsignificant findings. The significant findings reflect differences in performance on each question based on either methodological aspects of the way selective learning is investigated or differences in children's social cognitive processes. Given that the three questions show different relations with age, we considered only main effect models at first, and considered interaction models if those main effect models were significant.

The nonsignificant findings can be as illuminating as the significant ones; these results suggest what variables are unimportant to consider in evaluating empirical work. Moreover, these null results are important because they provide some control to the analysis we present. For example, there is no reason certain factors, like the year the article was published, should predict performance. Indeed, investigation of Tables 2–4 shows that this variable was not a significant predictor of performance on any of the three question types. Additionally, there are some cases where the failure to find a significant effect on any question potentially indicates that in the aggregate such a factor is not particularly impactful, even though individual publications might show varied results. Where necessary, we will temper our interpretation of the nonsignificant findings given concerns about power.

Sample Size

Our analysis uses age group/condition as the unit of analysis, and not experiment. The analysis focuses on a logit measure of performance, which does not include the size of the sample in the analysis. We thus examined whether sample size of the condition related to performance on any of the questions. Tables 2–4 show this was not the case. An inference we can make from this set of nonsignificant findings is that overall, the field of selective word learning used sufficiently powered findings. This suggests that the general effects of selective word learning are robust and replicable.

Informant Type

Responses to all three questions were not influenced by the nature of the informant and how the

Table 2
Summary of Analysis Results for Explicit Judgment Questions

Variable	Main effect
Nonsignificant	
Year of publication	$R^2 < .001$, $F(1, 172) = 0.19$, $p = .89$
Sample size of condition	$R^2 < .001$, $F(1, 172) = 0.04$, $p = .84$
Informant type (photo, video, live, puppet)	$R^2 = .016$, $F(3, 170) = 0.95$, $p = .42$
During familiarization only with multiple informants, was the information they presented in conflict?	$R^2 = .011$, $F(1, 168) = 1.94$, $p = .17$
At test, whether two informants label the same object with conflicting labels or there was only one informant generating a single label	$R^2 = .016$, $F(1, 172) = 2.80$, $p = .09$
When explicit judgment question was asked (before test, after test, both before and after, in the middle of test questions)	$R^2 = .025$, $F(3, 170) = 1.46$, $p = .23$

Variable	Main effect	Effect size
Significant		
Age (in months)	$R^2 = .028$, $F(1, 172) = 4.88$, $p = .028$	OR = 1.18
Number of informants during familiarization (one or two)	$R^2 = .045$, $F(1, 172) = 8.05$, $p = .005$	OR = 1.23
If explicit judgment question was asked, whether it was used as an inclusion criterion	$R^2 = .051$, $F(1, 169) = 9.03$, $p = .003$	OR = 1.25
Information generated during familiarization (epistemic, nonepistemic, both in conflict, both not in conflict)	$R^2 = .237$, $F(3, 180) = 17.57$ $p < .001$	$\eta_p^2 = .236$

Table 3
Summary of Analysis Results for Ask Questions

Variable	Main effect
Nonsignificant	
Year of publication	$R^2 = .002$, $F(1, 144) = 0.22$, $p = .64$
Sample size of condition	$R^2 = .001$, $F(1, 144) = 0.21$, $p = .65$
Informant type (photo, video, live, puppet)	$R^2 = .037$, $F(3, 142) = 1.83$, $p = .14$
Number of informants during familiarization (one or two)	$R^2 = .001$, $F(1, 144) = 0.15$, $p = .70$
During familiarization only with multiple informants, was the information they presented in conflict?	$R^2 = .015$, $F(1, 140) = 2.10$, $p = .15$
At test, whether two informants label the same object with conflicting labels or there was only one informant generating a single label	$R^2 = .003$, $F(1, 144) = 0.41$, $p = .52$
Whether explicit judgment question is asked	$R^2 = .011$, $F(1, 144) = 1.63$, $p = .20$
When explicit judgment question was asked, when was it asked (before test, after test, both before and after, in the middle of test questions)	$R^2 = .020$, $F(3, 106) = 0.72$, $p = .54$
When explicit judgment question was asked, whether it was used as an inclusion criterion	$R^2 = .001$, $F(1, 108) = 0.15$, $p = .70$

Variable	Main effect	Effect size
Significant		
Age (in months)	$R^2 = .081$, $F(1, 144) = 12.74$, $p < .001$	OR = 1.32
Information generated during familiarization (epistemic, nonepistemic, both in conflict, both not in conflict)	$R^2 = .052$, $F(3, 142) = 2.60$ $p = .05$	$\eta_p^2 = .052$

Table 4
Summary of Analysis Results for Endorse Questions

Variable	Main effect
Nonsignificant	
Age (in months)	$R^2 = .004$, $F(1, 344) = 1.22$, $p = .27$
Year of publication	$R^2 < .001$, $F(1, 344) = 0.06$, $p = .81$
Sample size of condition	$R^2 = .003$, $F(1, 344) = 1.16$, $p = .28$
Informant type	$R^2 = .002$, $F(3, 342) = 0.23$, $p = .87$
Number of informants during familiarization (one or two)	$R^2 = .009$, $F(1, 344) = 3.03$, $p = .08$
During familiarization only with multiple informants, was the information they presented in conflict?	$R^2 = .002$, $F(1, 276) = 0.48$, $p = .49$
Whether explicit judgment question is asked	$R^2 < .001$, $F(1, 344) = 0.06$, $p = .80$
When explicit judgment question was asked, when was it asked (before test questions, after test questions, both before and after, in the middle of test questions)	$R^2 = .011$, $F(3, 168) = 0.61$, $p = .61$
When explicit judgment question was asked, whether it was used as an inclusion criterion	$R^2 = .008$, $F(1, 164) = 1.39$, $p = .24$

Variable	Main effect	Effect size
Significant		
At test, whether two informants label the same object with conflicting labels or there was only one informant generating a single label	$R^2 = .012$, $F(1, 344) = 4.20$, $p = .04$	OR = 1.18
Information generated during familiarization (epistemic, nonepistemic, both in conflict, both not in conflict)	$R^2 = .035$, $F(3, 342) = 4.08$ $p = .007$	$\eta_p^2 = .034$

informant was presented to them (i.e., via puppet, live person, person on video, person in photograph). This null result might have been predictable, given that informant type was also not a significant predictor in children's false belief reasoning (Wellman et al., 2001). But beyond extending this result to the selective word learning paradigm, this result is interesting for two other reasons. First,

there has been methodological variability in how selective word learning procedures (and selective learning procedures more generally) are conducted. Using video to present familiarization and test events ensures consistency across the participants, but lacks a certain ecological validity. Using live informants is more ecologically valid, but introduces variability in the presentation of information. Live informants also introduce particular pragmatic demands to the procedure, as children must often reject an informant in their presence. Puppets often have the same variability as live informants, but potentially reduce those demand characteristics. Using puppets, however, has relatively low ecological validity. Critically, there do not appear to be many differences among these procedure types.

Second, there is a long literature suggesting that children are sensitive to others' pedagogy and social cues when engaged in word learning (e.g., Csibra & Gergely, 2009) as well as a general sensitivity for tracking reliability of information from social stimuli over nonsocial stimuli (e.g., Roseberry, Hirsh-Pasek, & Golinkoff, 2014). Although such experiments tend to use videos as opposed to live informants to present social stimuli, the videos used in those studies critically depict social events that allow for contingent action and reactions on the part of the child. This might suggest that children learn better from informants who provide them with such ostensive signals that communicate pedagogical intent. However, this hypothesis is not supported in our analysis. The failure to find such a result here is potentially consistent with literature that suggests children are also engaging in word learning through overhearing (e.g., Akhtar, Jipson, & Callanan, 2001). It is possible that mechanisms for selective word learning in young children might benefit from the presence of these pedagogical cues, but those cues might not be a necessary condition for selective learning to occur.

That said, this finding must be interpreted with caution. Unlike many of the other analyses presented here, in some cases, there were relatively few observations of some conditions, making these analyses not equally powered. For the Endorse question, this was not an issue, as there were at least 49 conditions for each type of informant. But for the Ask and Explicit Judgment question, there were relatively few cases of using live informants ($N = 7$ and $N = 3$, respectively). Coincidentally, that group seemed to be the worst performers among the four groups. It is possible that using live informants makes these questions more difficult for children, but we cannot come to that conclusion from this analysis.

Conflict Among Informants During Familiarization

An important demand characteristic of selective word learning procedures is the extent to which children observe conflicting information during familiarization. Early research on tracking epistemic competence emphasized preschoolers' capacities for selective word learning (e.g., Harris & Koenig, 2006; Koenig & Harris, 2005; Koenig et al., 2004). Investigation of younger children showed much more mixed results, with some suggesting 2-year-olds (for example) were capable of tracking others' epistemic competence and using that information for word learning (e.g., Brooker & Poulin-Dubois, 2013; Koenig & Woodward, 2010), and others suggesting that they did not have such capacities (e.g., Krogh-Jespersen & Echols, 2012). Koenig and Woodward (2010) suggested that a potential difference in the ways in which these studies are implemented might affect performance. The hypothesis was that toddlers could engage in selective word learning when there was conflict between informants during familiarization, but not when no conflict was present. Vanderbilt et al. (2014) systematically investigated the role of different kinds of conflict in preschoolers' selective word learning, suggesting places where it did and did not affect performance.

One Versus Two Informants

What is critical about this hypothesis is to define exactly what is meant by conflict. An initial difference between some of the Koenig and Woodward (2010) studies and the procedure used by Krogh-Jespersen and Echols (2012) is the number of informants with whom the children interact. Krogh-Jespersen and Echols (2012) suggested that when children only interacted with a single informant during familiarization, it might be hard for them to inhibit information generated by that informant, even if they demonstrated inaccurate information in the past because children had little other information upon which to base a decision. The demand characteristics of interacting with only one person might make rejecting the information presented by that single informant difficult, particularly for younger children.

The number of informants did not show a main effect on the Ask or Endorse question (although it was a marginal trend on the Endorse question).

Only on the Explicit Judgment question was there a significant effect of the number of informants. When researchers ask the Explicit Judgment questions, they rarely used a procedure with a single informant (only 4 conditions did so, contrasted with 170 that did not). The significant finding might reflect a ceiling effect, as on these four conditions, children averaged 92% accuracy (contrasted with 74% accuracy on the other conditions). It was also the case that researchers rarely asked the Ask question when there was only a single informant (also only 4 conditions), so we did not investigate this question further. We focus only on the Endorse question in the following section.

We analyzed whether performance on the Endorse question was affected by the interaction between the number of informants during familiarization and the age of the sample tested. The interaction between number of informants and age was significant, $R^2 = .015$, $F(1, 344) = 5.12$ $p = .024$, OR = 1.12. Overall, even the youngest children (2- to 3-year-olds) tested were better than chance on procedures in which they tracked a single informant (mean logit = .7303), whereas they were closer to chance when presented with multiple informants (mean logit = .4470). By age 4 and above, however, performance was mostly equivalent regardless of whether children observed 1 or 2 informants. The number of informants might affect younger children on the endorse question, but by the age of 4, this difference is not present.

Informants Generating Conflicting Information During Familiarization

We next considered how children performed when familiarized with more than one informant (or groups of informants). In some cases, children might observe two informants generate conflicting information about the same object (e.g., when shown a toy car, one might label it a "car" and the other label it a "shirt"). In others, they might see the two informants presenting different pieces of information that did not contain an explicit contrast. For none of the questions, however, did this factor show a main effect with performance.

To investigate this relation further, we looked at the interaction between age and conflict for all three question types. We found significant effects of this Age × Conflict interaction for the Ask question, $R^2 = .071$, $F(1, 140) = 10.64$ $p = .001$, OR = 1.30, and the Explicit Judgment question, $R^2 = .046$, $F(1, 168) = 8.16$, $p = .005$, OR = 1.24, but not the Endorse question. On the Ask question, children younger than 4 responded near chance levels when there was no conflict, but were consistently above chance when there was conflict. On the Explicit Judgment questions, children were also consistently above chance when there was conflict. When there was not conflict, children were mostly above chance as well, except for 5-year-olds, who showed poorer performance.

These interactions must be interpreted with caution, as in both cases, only 11 conditions in the analysis were cases where there was no conflict between the informants. The lack of significant main effects conflicts with findings by Vanderbilt et al. (2014), who suggested that the presence of this kind of conflict improved performance. The significant interaction on the Ask question, however, is consistent with Vanderbilt et al.'s findings, who limited their investigation to these relatively younger children. The nonlinear pattern of performance on the Explicit Judgment question, in contrast, is more difficult to interpret, but might be the result of the small sample. In general, our analyses suggest that this conflict matters for more the questions that require an inference to be made from others' reliability, but not for the demonstrative appreciation of others' information, as measured by the Endorse question.

What Information is Presented at Test

A third form of conflict involves whether there was conflict presented during the test phase or whether only a single informant generated information. For example, for an Endorse question, children could be shown a novel object and hear two different labels for it (usually by two different informants who have differed in some way in the familiarization phase of the procedure) or they could be shown that object and hear only one novel label (spoken by one of the informants). Hence, the conflict occurs at the test question, and not necessarily during the familiarization. Vanderbilt et al. (2014) investigated this scenario as well, and suggested that this did not affect performance. The results of the Ask and Explicit Judgment question are consistent with this finding, as this factor did not show a main effect on performance on these questions (although performance on the Explicit Judgment question was a marginal trend). As can be seen in Table 3, however, this factor did show a main effect for performance on the Endorse question.

To examine this effect further, we again examined the interaction between this factor and age, which showed a significant interaction on the

Endorse questions, $R^2 = .023$, $F(1, 344) = 8.09$, $p = .005$, OR = 1.16. The pattern here was similar to the previous analyses: When there is conflict, children younger than 4 responded near chance-like levels, whereas children over 4 were clearly above chance. When there is a single informant, all age groups were above chance levels.

Summary of Analysis on Informant Conflict

Taken together, these data suggest that conflict potentially has distinct effects on responses to the different types of questions. Having multiple informants, and specifically having those informants conflict on the information presented during the test trial affects the developmental trajectory of responses to the Endorse question. Conflict during the training by multiple informants affects the developmental trajectory of responses to the Ask question. In both cases, it does appear that around age 4, there are some changes in the way children respond to these factors, with children older than 4 mostly performing quite well, whereas children younger than 4 were closer to chance in their performance.

Given these findings, we want to revisit the results of Vanderbilt et al. (2014). In their studies, they found that the presence of two informants who generated conflicting information during familiarization promoted selective learning for the accurate informant on an endorse question; when only one informant was present during familiarization, children relied on that informant even when she was inaccurate. This study was included in our analysis, and we do not doubt the findings presented there. The results of our analysis, however, suggests that looking across a larger data set and across numerous methods, children might engage in selective word learning when presented with either single or multiple informants presenting conflicting information. That said, in some cases, there were still only a small number of studies to consider, and further research specifically looking at the role of pragmatics inferences children must make when they observe conflicting information is needed.

Role of Explicit Judgment Question

Explicit Judgment questions often require children to make a general judgment about the valence of the informants as opposed to making an inference about a specific piece of knowledge or label for a novel object (e.g., "One of these people was not very good at answering these questions. Which one was not very good at answering questions?" - taken from Koenig & Harris, 2005, p. 1264). Some researchers use Explicit Judgment questions as control questions for the Ask and Endorse questions, and only include data from children who pass this question. Other studies treat this question as a separate data point, either analyzed in combination with or alongside other test questions. There is not a systematic use for these questions across studies.

Because Explicit Judgment questions involve making a more general inference about the informants' epistemic competence, they also have the potential to influence performance on the other test questions. Studies have varied when the Explicit Judgment question is asked in relation to other test questions. Studies that ask this question prior to other test questions might bias children to respond to those test questions in the same manner as the Explicit Judgment question. Why choose to endorse one informant when the child has stated that the other is better at the game? This is particularly important if one uses the Explicit Judgment question as an inclusion criterion. Our analyses showed that while performance on the Explicit Judgment question itself improved with age, whether and when it was asked during the procedure as well as whether it was used as an inclusion criterion for analysis did not affect performance on either the Ask or Endorse questions. The only significant finding here of note is that when the explicit judgment question was used as an inclusion criterion, performance on it was better than when it was not. This, however, seems to be an artifact of using it as an inclusion criterion.

The null results here point out an important methodological point, which is that the explicit judgment question is not necessarily a control question. In their analysis of the unexpected contents task, Sobel and Austerweil (2016) found that performance on that measure differed depending on how responses to the control question were factored into the scoring. Here, performance on the explicit judgment question should not be considered a control, but rather tapping a different type of knowledge.

We suggest that a better control for selective learning measures was proposed indirectly by Sabbagh and Shafman (2009), who found that preschoolers were often accurate about their episodic memory for the labeling episode of ignorant informants, even if they did not necessarily learn the label for the object. That is, children could often answer a question about the labeling event (i.e., which one did she say was a "blicket") even if they

discounted the information from an ignorant informant (i.e., they did not refer to that object as a "blicket" in subsequent testing). Although Sabbagh and Shafman did not test this as a within-subject manipulation, questions about epistemic memory might serve as better controls for selective word learning (i.e., semantic memory acquisition) than explicit knowledge questions.

Finally, many studies that consider Endorse questions also asked Ask questions. These questions are usually asked one after the other—with the Ask question preceding the Endorse question. Responding to the Ask question might have influenced performance on the Endorse question, as stating that one would like to ask one informant might nudge children to pick that informant's information in response to the Endorse question. Of the 346 conditions included in our analysis where the Endorse question was asked, 154 asked an Ask question prior.

We constructed a regression analysis looking at performance on the Endorse question as the dependent measure with whether an Ask question was also asked and its interaction with age. The overall model was significant, $R^2 = .03$, $F(2, 343) = 4.94$, $p = .008$. Children were better at responding on the Endorse question when they were also asked an Ask question than children who were only asked an Endorse question, $B = -0.85$, $SE = 0.40$, $t = -2.13$, $p = .03$. Moreover, this relation interacted with age, with the effect more pronounced as children got older, $B = 0.02$, $SE = 0.007$, $t = 2.61$, $p = .009$. What these results suggest is while the Explicit Judgment question does not affect performance on the Ask or Endorse question, the presence of the Ask question might influence performance on the Endorse question, particular as children get older.

Epistemic Versus Nonepistemic Factors

The analyses so far have focused on methodological factors that potentially affect performance on studies of selective word learning. We now turn to a set of analyses that examine what cues children use. We categorized studies based on whether researchers presented children with informants who only demonstrated their epistemic competence (i.e., information about accuracy or expertise) or who provided cues that could be used as a basis for selective learning, but did not necessarily indicate epistemic competence (i.e., were judged to be less of a basis for selective learning by our coders). In many cases, coders judged that both epistemic and nonepistemic factors were present, and in those cases, we coded whether they were consistent with each other or in conflict. As shown in Tables 2–4, differences among these four categories predicted performance on all three question types. A breakdown of the overall proportion correct for these factors is shown in Table 5.

We also looked at whether there was a significant interaction between this category and age. This interaction was not significant for the Ask question, but was significant for the Endorse questions, $R^2 = .03$, $F(3, 342) = 3.53$, $p = .015$, and for the Explicit Judgment questions, $R^2 = .25$, $F(3, 170) = 19.07$, $p < .001$. On the Endorse question, there was a significant correlation between performance and age (in months) when only nonepistemic factors were manipulated, $R^2 = .14$, $F(1, 86) = 13.96$, $p < .001$, OR = 1.45, but no other significant correlation with age for any of the three other cases. On the Explicit Judgment question, there was only a significant correlation between performance and age when both epistemic and nonepistemic factors were present, but not in conflict, $R^2 = .182$, $F(1, 25) = 5.56$, $p = .026$, OR = 1.53, but in no other case.

These analyses suggest that children respond differently to manipulations in selective word learning that focus on establishing informants' epistemic competence as opposed to factors that might

Table 5
Overall Performance (Proportion Correct) on Test Questions Based on the Type of Information Generated by Informants

	Epistemic factor only	Nonepistemic factor only	Both factors in conflict	Both factors not in conflict
Ask question	.67	.69	.60	.81
	$N = 100$	$N = 31$	$N = 7$	$N = 5$
Endorse question	.67	.62	.55	.77
	$N = 228$	$N = 88$	$N = 18$	$N = 12$
Epistemic judgment question	.78	.64	.60	.86
	$N = 104$	$N = 13$	$N = 30$	$N = 27$

Note.. Performance was defined as appealing to the informant who was (more) epistemically competent, who had (more) positive valence, or who was (more) in line with the hypotheses of the particular study.

influence children's beliefs that others are reliable sources of knowledge. Not only do children respond judiciously among studies that manipulate different types of information about the informants, but they respond differently among the three question types. Children's age also affects performance differently across these questions.

What these analyses do not tell us is whether there are differences among the nonepistemic factors that have been investigated. Table 6 shows the breakdown of these factors and the results of this analysis. We used a sample size of $N = 20$ conditions as a criterion to analyze these data, as any sample smaller than that would potentially have insufficient power. There was a relation between performance and age on the Ask and Endorse questions when the type of informant was manipulated, but not for the Explicit Judgment question. Similarly, when linguistic factors were manipulated, older children did better on the Endorse questions, but there was not sufficient power to examine the Ask or Explicit Judgment questions. In contrast, there was no relation with age and consensus, suggesting children's sensitivity to consensus was consistent throughout the ages we considered. A limitation, however, is that the above analysis summarizes over cases in which the nonepistemic factor was manipulated alone or with an epistemic factor like accuracy, either consistently or in conflict. Breaking the analysis down any further reduces the sample size sufficiently to raise concerns about the power of the analyses.

Discussion

Our analysis focused on children's selective word learning. We summarize the main findings of the analysis as follows: (a) Question type matters; (b) Certain methodological factors matter; (c) Different cues to selective learning matter. We will discuss each of these findings in the following section, then integrate our findings with a published set of meta-analyses (Tong et al., 2019), describe the implications of this analysis on mechanisms that might underlie selective word learning, as well as present limitations and future directions based on this analysis.

Question Type Matters

Performance on selective word learning measures depends on what kind of question is posed to children. Performance on the Ask, Endorse, and Explicit Judgment questions differ because what these questions are asking of children differs. The endorse question, which was most commonly asked in our analysis, focuses on whether children use the information generated by an informant. The Explicit Judgment and Ask questions focus on more inferential facets of selective learning. The former is about valence—who is better or worse at generating labels or who is more or less knowledgeable? The latter is more metacognitive—whom should be queried in order to obtain knowledge.

Performance on the Endorse question did not correlate with age, whereas performance on the

Table 6
Number of Conditions That Examined Each Nonepistemic Factor by Question and Relations With Age for Qualifying Power Criterion ($N \geq 20$*)*

	Ask questions	Endorse questions	Explicit judgment questions
Linguistic factors	$N = 0$ did not meet power criterion	$N = 21$ $R^2 = .29$ $F(1, 19) = 7.71$ $p = .01$	$N = 16$ did not meet power criterion
Perceptual valence	$N = 7$ Did not meet power criterion	$N = 12$ Did not meet power criterion	$N = 1$ Did not meet power criterion
Type of informant	$N = 29$ $R^2 = .22$ $F(1, 27) = 7.44$ $p = .01$	$N = 49$ $R^2 = .21$ $F(1, 47) = 12.57$ $p = .001$	$N = 49$ $R^2 = .04$ $F(1, 47) = 2.19$ $p = .15$
Consensus	$N = 10$ Did not meet power criterion	$N = 26$ $R^2 = .04$ $F(1, 24) = 0.92$ $p = .35$	$N = 4$ Did not meet power criterion
Other	$N = 0$ Did not meet power criterion	$N = 12$ Did not meet power criterion	$N = 2$ Did not meet power criterion

Explicit Judgment and Ask questions did. Performance on these questions also differed across ages, as indicated by the different regression lines with age. The Ask question involves more of a developmental trajectory than the other two, suggesting that children might be developing more of a metacognitive awareness of others' reliability through the preschool years, but their use of others' information might be in place relatively early in development.

These differences offer a way of integrating the literature on selective word learning. Children potentially have the ability to make trait-like judgments about others' reliability early in development in the form of explicit judgment questions, although this capacity improves with age. What children seem to lack is the capacity to make metacognitive inferences early in development, consistent with the idea that they are initially credulous and not evaluating others. Children can make inferences about the valence of an informant—whether they are good or bad early in development—but do not use that valence to decide whom to query for information.

But what children can do early in development is use information from others judiciously as shown by our analysis of the Endorse questions. This does not mean that children always do so; there are situations where children need other cognitive capacities that have independent developmental trajectories in order to learn selectively (such as inhibitory control, e.g., Jaswal et al., 2014, or knowledge access, Nurmsoo & Robinson, 2009). This suggests that children have a general capacity for tracking others that allows them to register whose information to use, but specific knowledge of how to evaluate that information has its own developmental trajectories, which might explain why children sometimes learn selectively and sometimes do not.

Finally, children's different performance on the Ask and Endorse questions affords potentially different interpretations of what these two questions reveal. The Ask question allows children two alternatives in their responses: they could exploit their understanding of the epistemic competence of the informants or explore what the informants are going to say in order to learn more about them? Correctly responding on the Ask question could be seen as children exploiting the knowledge that they already possess about the informants. Incorrectly responding could be seen as children behaving inaccurately or as them exploring the informant whom they believe is less epistemically reliable. Outside of selective word learning, Turner, Giraldeau, and Flynn (2017) showed that children paid attention to unreliable sources of information, which suggests that responses on the Ask question might indicate that children know they are asking the inappropriate informant, but do so because it is salient to figure out why that person has been inaccurate. This is, however, an open question for further investigation.

Methodology Matters

A methodological recommendation based on this analysis is to not combine the results of these questions into a single metric (typically, the Ask and Endorse questions), as has been done in many studies, as these questions are not eliciting the same type of knowledge. Moreover, a recommendation is not to even ask both types of questions to the same participant. We saw that asking an Ask question influenced performance on the Endorse question, particularly as children got older. Children might find it awkward to ask one informant for information, but then use the information generated by the other.

In contrast, while there seemed to be a clear influence of the Ask on the Endorse question, what did not seem to matter much to children's performance on either question was whether an Explicit Judgment question was asked. This suggests that the way in which children are responding to this question is independent to the way they are making inferences about the Ask or the Endorse questions. Making an inference about whether an informant is a good or bad source of knowledge might be independent of choosing whether to use that informant's information or whether that informant has a particular piece of knowledge.

Finally, different aspects of conflict among informants during familiarization selectively affected performance on individual test questions. These findings mostly interacted with age, making younger children more susceptible to these methodological differences. For example, younger children were more affected when two informants conflicted in the information they generated. Such conflict might allow younger children to think through the contrast between the informants' information. This hypothesis is related to the literature on analogical reasoning. For example, Christie and Gentner (2010) showed that 3-year-olds were more likely to make a relational inference about the extension of a novel label when given comparison data than when given one of those data points individually. In this study,

the comparison data serve to highlight the relation as opposed to other features that children could use to make the inference. Conflict might highlight that the different information indicates distinct levels of epistemic competence, which might be required for younger children to make the inference. Critically, this conflict affected only the Endorse questions—the decision to use the information generated by others, and not more metacognitive or trait-like inferences about others. This is consistent with the possibility that the decision to use others information is independent from judgments about those informants; the information is independent of the person generating it.

Cues Matter

We investigated the role that different cues presented to children had in their judgments. We divided cues into those that necessarily indicated epistemic competence (such as accuracy or competence) and those that could, but did not necessarily do so (such as social in-group or consensus). These cues are not mutually exclusive. When both cues were manipulated such that there was no conflict between them, children were more likely to use that information than when either cue was manipulated alone or when those two cues conflicted. The latter resulted in the lowest levels of selective learning (see Table 5). Epistemic and nonepistemic cues on their own seemed to be equally effective in promoting selective learning.

We wish to make two points about these findings. The first is to consider that even in experiments where epistemic and nonepistemic cues are in conflict, children tend to favor epistemic cues. The nonepistemic cues, however, are potentially more heterogeneous in their relevance for selective word learning whereas the epistemic cues might be more homogeneous in what they indicate about others' reliability. It is certainly possible that some nonepistemic cues are more heavily weighted as cues to selective learning than others, and even more heavily weighted than epistemic cues. This is, however, an empirical question. It might also be the case that as children develop—as their domain-specific knowledge about these nonepistemic cues or about knowledge more generally (i.e., their metacognition) changes—they value these cues differently, particularly in comparison with epistemic cues (Sobel & Kushnir, 2013).

Second, that epistemic and nonepistemic cues individually result in similar levels of selective learning calls back the distinction between selective learning and selective trust. Harris et al. (2018) argue that there is more to selective learning than making an inference about the epistemic competence of others. Here, and elsewhere (e.g., Jaswal & Kondrad, 2016), they suggest that a broader theoretical claim is necessary about the social relationships that children have with others. Selective *trust* is thus a broader conceptualization of selective learning. Judgments of epistemic competence is one facet of such trust, but so are what they call "normative" aspects—for example, that we communicate with each other via speech acts is a sign of trust, "as well as the greater stock of reasons that children, and all agents, have to trust another person" (Harris et al., p. 254). Thus, cues about others' epistemic competence might be part of the mechanism children use for selective learning. Cues that do not necessarily point to epistemic competence (e.g., perceptual information about the informant) are important because they indicate social relationships that children have with others—all part of "greater stock" that children use as their basis for trust. Thus, in some cases, children are responding to these questions because they make an inference about the epistemic states of the informants, but in others, the respond because of a more nonepistemic trust. What is necessary to move this discussion forward is to articulate what kinds epistemic and nonepistemic cues go into interpersonal trust as opposed to judgments of epistemic competence.

Relations to Tong et al. (2019) Meta-Analyses

We also want to emphasize similarities and differences between the present analysis and those reported by Tong et al. (2019), as well as discuss why differences might be present. First, regarding the role of age in children's judgments, our findings on the Ask question are generally congruent with Tong et al. They found a significant effect of age only on the Ask questions when only epistemic cues are manipulated (their Analysis 1). The interaction between age and manipulated factors was not significant in our analysis for the Ask question. However, if we look at performance on the Ask questions in this circumstance, we do find a marginally significant relation with age, $r(116) = .17$, $p = .07$. While potentially Type I error, this analysis coupled with the absence of effects of age on other Ask questions is consistent with the Tong et al. finding.

Where the results of this analysis differ from the results of Tong et al is on the Endorse question. Here, children did not show a significant effect of

age on the Endorse questions. But the present analysis considers selective word learning experiments together, while Tong et al. divides their analysis into three categories—manipulations of only epistemic cues, manipulations of only social characteristics (labeled here nonepistemic cues), and manipulations where these cues are in conflict. To parallel their analysis, we can look at the interaction between what cues are presented to children and age. Unlike Tong et al., we did not find a significant correlation with age and performance on the Endorse question when only epistemic cues were considered (akin to their Analysis 1, where the effect of age was significant) or when epistemic and nonepistemic cues were in conflict (their Analysis 3 where the effect of age was marginally significant), but we did find a significant correlation with age when only nonepistemic factors were manipulated (their Analysis 2 where the effect of age was not significant).

There are several potential reasons for these differences as well as the similarities in the effect of age on the ask question. The first is that we did include several studies that examined children younger than age 3 (e.g., Brooker & Poulin-Dubois, 2013; Koenig & Woodward, 2010; Krogh-Jespersen & Echols, 2012; Luchkina et al., 2018) and a more expansive age range might affect these results. Studies on children younger than 3 rarely ask the ask question. Our analysis of the Ask question considered samples with mean ages between 40–-84 months, whereas our analysis of the Endorse question considered samples with mean ages between 18–92 months. Second, treating age continuously might be a more sensitive way of analyzing age than grouping children by year. This can be seen in the intercept values of the Ask, Endorse, and Explicit Judgment questions, which suggest that the ask question is a harder question overall.

The other important difference is potentially more theoretical. We separated out different nonepistemic factors to attempt to analyze them individually. Rarely were we able to achieve a criterion for sufficient power in these analyses, but where we could, we found that nonepistemic cues that might convey certain pieces of epistemic information (such as linguistic factors like accent or type of informant like whether an adult or child generated information) did seem to show correlations between age and performance on the endorse questions. In contrast, consensus did not. The former two might convey more information about an informant's epistemic competence than the latter, and older children might be more sensitive to this possibility.

Psychological Mechanisms of Selective Word Learning

We have suggested that the three questions typically used to study selective word learning (Ask, Endorse, Explicit Judgment) potentially test different facets of children's understanding of others' epistemic competence. An open question is what this analysis can tell us about the mechanisms that underlie how children respond to these questions and more generally make judgments about others' epistemic competence.

To consider this question, it is important to enumerate what mechanism might affect children's judgments. Kuzyk, Grossman, and Poulin-Dubois (2019) categorize mechanisms into those that are more domain-general or more domain-specific. Heyes (2017), for example, suggests that a particular domain-general mechanism—associative learning—can be applied to phenomena in selective learning. Children have associative learning mechanisms available to them in infancy (Kirkham, Slemmer, & Johnson, 2002) and infants can make selective inferences to guide their attention based on such mechanisms (e.g., Tummeltshammer, Wu, Sobel, & Kirkham, 2014). By learning associations between speaker-label-object pairings, children might come to recognize that some informants are more epistemically competent over others based on the strength of those associations. Similarly, Jaswal et al. (2014) demonstrated a relation between children's trust in (inaccurate) verbal testimony as opposed to their first-hand experience and their performance on inhibitory control measures. Inhibiting false statements from others potentially requires domain-general cognitive control capacities (Jaswal & Kondrad, 2016).

The lack of a significant relation between age and performance on the Endorse question speak to these mechanisms. Given the substantial development of inhibitory control during early childhood (e.g., Davidson, Amso, Anderson, & Diamond, 2006) it seems unlikely that inhibitory control alone influences selective word learning. That very young children seem to endorse others' information based on epistemic competence does suggest the possibility that associative learning influences some facet of children's judgments about selective learning, as such mechanisms seem to be available to these very young children. However, other findings, such as the relations with age for the Ask and Explicit Judgment questions, are inconsistent with the

hypothesis that such domain-general mechanisms alone account for selective word learning.

On the domain-specific side, many researchers have articulated correlations between children's selective learning and their performance on the false belief task (e.g., Brosseau-Liard, Penney, & Poulin-Dubois, 2015; DiYanni, Nini, Rheel, & Livelli, 2012). Others, however, have not found such relations (e.g., Pasquini, Corriveau, Koenig, & Harris, 2007) or found only limited relations (Van Reet, Green, & Sobel, 2015). Similar to the arguments for children's developing inhibitory control, the present findings do not necessarily support the hypothesis that false belief knowledge underlies children's selective learning. Of import, however, is that the Brosseau-Liard et al. (2015) investigation administered the ToM scales (Wellman & Liu, 2004), as opposed to the other studies that looked only at false belief performance. It is possible that the ToM scales are more indicative of children's general mental state knowledge as opposed to whether others can have false beliefs about the world. Indeed, Crivello, Phillips, and Poulin-Dubois (2017) suggested that toddlers' ability to make inferences about the knowledge states of others related to their selective learning and Kuzyk et al. (2019) found that toddlers' developing metacognition (their ability to reflect on their own thinking) predicted whether they learned from unreliable informants. This suggests that possessing domain-specific knowledge about others' knowledge states more generally might underlie certain aspects of children's selective learning.

A compromise between these views comes from thinking about mechanisms that posit both low-level domain-general associative reasoning and higher level domain-specific cognitive processes as underlying children's selective learning (e.g., Hermes et al., 2016; Sobel & Kushnir, 2013). That different selective word learning questions have different developmental trajectories and different degrees of difficulty are consistent with this possibility. Luchkina, Morgan, and Sobel (in press) demonstrated that preschoolers' selective word learning is explained by domain-specific cognitive capacities, but certain influences of domain-general associative learning mechanisms underlie some facets of children's responses. Given that children register information from knowledgeable and ignorant speakers differently at a neural level (Mangardich & Sabbagh, 2018), children might attend to both lower level domain general and higher-level domain-specific information when making selective judgments. Articulating the nature of these two mechanisms (potentially with computational modeling to specify these algorithms more precisely, see Sobel, Buchanan, Butterfield, & Jenkins, 2010, for one example) is a topic for future investigation.

Limitations of the Present Analysis

While our discussion of theory has focused on selective learning generally construed, it is important to remember that our analysis is specifically on selective word learning, and we should take care to restrict our proximal interpretations to only this domain. We limited our analysis to selective word learning for both practical and theoretical reasons. Many of the questions in selective imitation or other selective learning literatures have open-ended responses, making a direct comparison to the two alternative forced-choice questions asked in selective word learning difficult. Moreover, unlike selective imitation, selective word learning might be a uniquely human capacity and governed by a uniquely human mechanism, making a summary of the selective learning literatures *in toto* potentially less ideal than an analysis of each individually. It is possible that the present results will not generalize to other forms of selective inference.

While this is an empirical question, we want to suggest that the results would generalize broadly to these other literatures. In human development, precursors to selective learning—like the ability to learn from pedagogical cues (Csibra & Gergely, 2009)—focus on our communicative nature. As Heyes (2017) points out, even lower animals learn selectively, suggesting that such a capacity might have an evolutionary basis. Indeed, relatively young infants are sensitive to situations when adults are trying to communicate with them (e.g., Vouloumanos & Werker, 2004, 2007), and have relatively sophisticated capacities for parsing meaningful communicative utterances from others' sounds (e.g., Perszyk, Ferguson, & Waxman, 2018). Moreover, in some of the direct comparisons between selective word learning and selective imitation, children seem to have similar capacities (e.g., Brooker & Poulin-Dubois, 2013). These findings all suggest the possibility that selective learning more generally construed might be the basis for the more sophisticated cultural transmission processes that govern interaction. What is necessary is a similar analysis to what we have presented here on the selective imitation literature, but this is a goal for future investigations.

Second, an important limitation of our analysis comes from considering how different nonepistemic

cues influence selective learning. We used an omnibus category, which contrasted cues that clearly showed epistemic competence with cues that did not necessarily do so. However, this latter group was comprised of cues that were potentially more related to a continuum than a strict category. The age of the informant, for example, could be a good clue to epistemic competence, as adults tend to know more than children. But, as pointed out by VanderBorght and Jaswal (2009), there are some domains of knowledge (e.g., the names of Pokemon characters) where children's knowledge often exceeds that of adults'. Indeed, we found some places where sensitivity to nonepistemic cues (at least for the Endorse questions) increases with age (such as linguistic factors and the type of informant), whereas other places where there was no relation to age (such as consensus of information). However, care must be taken in interpreting these data, as combining nonepistemic factors together does not suggest that children interpret each of these factors via the same cognitive mechanism.

Third, we want to point out that focusing on the type of question children are asked might limit the analysis to the methods used to study selective learning, as opposed to focusing on the cues that children use to learn from others. However, given that children respond to the three questions differently, it is important to refocus investigations on selective learning by considering what cues affect each cognitive process represented by the different questions. This is potentially a best practice for moving the field forward, as opposed to simply combining responses from the questions together and treating that measure as a singular cognitive process.

Finally, like similar analyses in ToM (Sobel & Austerweil, 2016; Wellman & Liu, 2004; Wellman et al., 2001), the present analysis focuses on the proportion of appropriate performance (as we have defined it earlier) across age group and condition. Our unit of analysis was not publication or experiment, but rather conditions within an experiment based on age groups. An advantage of this approach is that we were able to consider certain contrasts (such as single informant procedures vs. multiple informant procedures) that other approaches cannot consider. A limitation is that our analysis weighs conditions with different sample sizes equivalently. Given that we find some similar results to Tong et al. (2019), who take a different approach in their analysis, and where there are differences might result from particular inclusion criteria across the two analyses, we think these methods provide complementary conclusions.

Conclusion

Children learn from others, track the accuracy of others' information, and discount potentially inaccurate, irrelevant, or uninformative sources. Our goal was to synthesize both disparate methods and different theoretical perspectives. Methodologically, the analysis suggests that some factors are relevant to performance. Different questions related to selective learning tap into different cognitive capacities and have different developmental trajectories. Theoretically, children use different cues to make selective inferences, and while their inferences about others information and others as sources of information might differ, their inferences about learning from others might be integrated with their inferences about whether to trust others.

While it is possible that there is a single, unified theory of selective learning that would more parsimoniously explain all these findings, we suspect that the differential nature of the test questions used in these studies—questions that span children's cognitive and metacognitive capacities—indicates that numerous developing processes and pieces of knowledge are necessary for the complexity of children's selective learning. Some of these questions are more about the rational processing of information; others are more about the social relationship children might have with informants, and both these factors might contribute to what makes children such remarkable learners.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*, 416–430. https://doi.org/10.1111/1467-8624.00287

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1–18. https://doi.org/10.1037/a0021017

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. https://doi.org/10.1111/bjdp.12022

Bascandziev, I., & Harris, P. L. (2016). The beautiful and the accurate: Are children's selective trust decisions biased? *Journal of Experimental Child Psychology*, *152*, 92–105. https://doi.org/10.1016/j.jecp.2016.06.017

Bergstrom, B., Moehlmann, B., & Boyer, P. (2006). Extending the testimony problem: Evaluating the truth, scope, and source of cultural information. *Child Development*,

77, 531–538. https://doi.org/10.1111/j.1467-8624.2006.00888.x

Bernard, S., Proust, J., & Clément, F. (2015). Four- to six-year-old children's sensitivity to reliability versus consensus in the endorsement of object labels. *Child Development*, *86*, 1112–1124. https://doi.org/10.1111/cdev.12366

Bernath, M. S., & Feshbach, N. D. (1995). Children's trust: Theory, assessment, development, and research directions. *Applied and Preventive Psychology*, *4*(1), 1–19. https://doi.org/10.1016/S0962-1849(05)80048-4

Birch, S. A. J., & Bloom, P. (2002). Preschoolers are sensitive to the speaker's knowledge when learning proper names. *Child Development*, *73*, 434–444. https://doi.org/10.1111/1467-8624.00416

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2007). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*, 1018–1034. https://doi.org/10.1016/j.cognition.2007.12.008

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

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*, 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (1998). Explanation in scientists and children. *Minds and Machines*, *8*, 119–136. https://doi.org/10.1023/A:1008242619231

Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, *18*, E46–E68. https://doi.org/10.1111/infa.12027

Brosseau-Liard, P., Cassels, T., & Birch, S. (2014). You seem certain but you were wrong before: Developmental change in preschoolers' relative trust in accurate versus confident speakers. *PLoS One*, *9*, e0108308. https://doi.org/10.1371/journal.pone.0108308

Brosseau-Liard, P., Penney, D., & Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *British Journal of Developmental Psychology*, *33*, 464–475. https://doi.org/10.1111/bjdp.12107

Canfield, C. F., Saudino, K. J., & Ganea, P. A. (2015). The role of temperament in children's reliance on others as sources of information. *Infant and Child Development*, *24*, 435–451. https://doi.org/10.1002/icd.1892

Chen, E. E., Harris, P. L., & Corriveau, K. H. (2012). Children trust a consensus composed of outgroup members—But do not retain that trust. *Child Development*, *84*, 269–282. https://doi.org/10.1111/j.1467-8624.2012.01850.x

Chi, M. T., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477. https://doi.org/10.1016/0364-0213(94)90016-7

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*, 356–373. https://doi.org/10.1080/15248371003700015

Clément, F., Bernard, S., Grandjean, D., & Sander, D. (2013). Emotional expression and vocabulary learning in adults and children. *Cognition & Emotion*, *27*, 539–548. https://doi.org/10.1080/02699931.2012.724012

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*, 360–379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford, UK: Clarendon Press.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *3*, 372–377. https://doi.org/10.1111/j.1467-9280.2009.02291.x

Corriveau, K. H., & Harris, P. L. (2009). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*, 188–193. https://doi.org/10.1111/j.1467-7687.2008.00763.x

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426–437. https://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. https://doi.org/10.1037/a0030604

Corriveau, K. H., & Kurkul, K. E. (2014). "Why does rain fall?": Children prefer to learn from an informant who uses noncircular explanations. *Child Development*, *85*, 1827–1835. https://doi.org/10.1111/cdev.12240

Corriveau, K. H., Kurkul, K. E., & Arunachalam, S. (2016). Preschoolers' preference for syntactic complexity varies by socioeconomic status. *Child Development*, *87*, 1529–1537. https://doi.org/10.1111/cdev.12553

Corriveau, K. H., Pickard, K., & Harris, P. L. (2011). Preschoolers trust particular informants when learning new names and new morphological forms. *British Journal of Developmental Psychology*, *29*(1), 46–63. https://doi.org/10.1348/2044-835X.002009

Crivello, C., Phillips, S., & Poulin-Dubois, D. (2017). Selective social learning in infancy: Looking formechanisms. *Developmental Science*, *21*, e1259. https://doi.org/10.1111/desc.12592

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*, 304–316. https://doi.org/10.1111/1467-8624.00023

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Danovitch, J. H., & Alzahabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, *14*, 499–513. https://doi.org/10.1080/15248372.2012.689391

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02

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

DeLeeuw, N., & Chi, M. T. H. (2003). Self-explanation: Enriching a situation model or repairing a domain model. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 55–78). Mahwah, NJ: Erlbaum.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, *71*, 329–338. https://doi.org/10.1111/1467-8624.00148

DiYanni, C., Nini, D., Rheel, W., & Livelli, A. (2012). "I won't trust you if I think you're trying to deceive me": Relations between selective trust, theory of mind, and imitation in early childhood. *Journal of Cognition and Development*, *13*, 354–371. 0.1080/15248372.2011.590462

Doebel, S., & Koenig, M. A. (2013). Children's use of moral behavior in selective trust: Discrimination versus learning. *Developmental Psychology*, *49*, 462–469. https://doi.org/10.1037/a0031595

Doebel, S., Koenig, M. A., & Rowell, S. F. (2016). Young children detect and avoid logically inconsistent sources: The importance of communicative context and executive function. *Child Development*, *87*, 1956–1970. https://doi.org/10.1111/cdev.12563

Einav, S., & Robinson, E. J. (2010). Children's sensitivity to error magnitude when evaluating informants. *Cognitive Development*, *25*, 218–232. https://doi.org/10.1016/j.cogdev.2010.04.002

Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science*, *22*, 1250–1253. https://doi.org/10.1177/0956797611416998

Elashi, F. B., & Mills, C. M. (2014). Do children trust based on group membership or prior accuracy? The role of novel group membership in children's trust decisions. *Journal of Experimental Child Psychology*, *128*, 88–104. https://doi.org/10.1016/j.jecp.2014.07.003

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. https://doi.org/10.1016/j.jecp.2011.06.008

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, *11*, 771–777. https://doi.org/10.1111/j.1467-7687.2008.00728.x

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Gopnik, A. (1998). Explanation as orgasm. *Minds and Machines*, *8*, 101–118. https://doi.org/10.1023/A:1008290415597

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*, 1085. https://doi.org/10.1037/a0028044

Guerrero, S., Elenbaas, L., Enesco, I., & Killen, M. (2017). Preschoolers' trust in social consensus varies by context: Conventional vs. moral domains. *Anales de Psicología*, *33*, 142–151. https://doi.org/10.6018/analesps.32.3.230831

Harris, P. L. (2002). What do children learn from testimony? In P. Carruthers, S. P. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 316–334). Cambridge, UK: Cambridge University Press.

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

Harris, P. L., & Koenig, M. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*, 505–524. https://doi.org/10.1111/j.1467-8624.2006.00886.x

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*, 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science*, *9*, 76–96. https://doi.org/10.1111/j.1467-7687.2005.00465.x

Hermes, J., Behne, T., Studte, K., Zeyen, A.-M., Gräfenhain, M., & Rakoczy, H. (2016). Selective cooperation in early childhood—How to choose models and partners. *PLoS One*, *11*, e0160881. https://doi.org/10.1371/journal.pone.0160881

Heyes, C. (2017). When does social learning become cultural learning? *Developmental Science*, *20*(2), 1–30. https://doi.org/10.1111/desc.12350

Jaffer, S., & Ma, L. (2015). Preschoolers show less trust in physically disabled or obese informants. *Frontiers in Psychology*, *5*, 1524. https://doi.org/10.3389/fpsyg.2014.01524

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology*, *61*, 248–272. https://doi.org/10.1016/j.cogpsych.2010.06.002

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*, 1541–1547. https://doi.org/10.1177/0956797610383438

Jaswal, V. K., & Kondrad, R. L. (2016). Why children are not always epistemically vigilant: Cognitive limits and social considerations. *Child Development Perspectives*, *10*, 240–244. https://doi.org/10.1111/cdep.12187

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. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Jaswal, V. K., Pérez-Edgar, K., Kondrad, R. L., Palmquist, C. M., Cole, C. A., & Cole, C. E. (2014). Can't stop believing: Inhibitory control and resistance to misleading testimony. *Developmental Science*, *17*, 965–976. https://doi.org/10.1111/desc.12187

Keil, F. C. (2010). The feasibility of folk science. *Cognitive Science*, *34*, 826–862. https://doi.org/10.1111/j.1551-6709.2010.01108.x

Keil, F. C., & Wilson, R. A. (Eds.). (2000). *Explanation and cognition*. Cambridge, MA: The MIT Press.

Kim, S., & Harris, P. L. (2014a). Belief in magic predicts children's selective trust in informants. *Journal of Cognition and Development*, *15*, 181–196. https://doi.org/10.1080/15248372.2012.751917

Kim, S., & Harris, P. L. (2014b). Children prefer to learn from mind-readers. *The British Psychological Society*, *32*, 375–387. https://doi.org/10.1111/bjdp.12044

Kim, S., Paulus, M., & Kalish, C. (2016). Young children's reliance on information from inaccurate informants. *Cognitive Science*, *41*(Suppl. 3), 601–621. https://doi.org/10.1111/cogs.12471

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42. https://doi.org/10.1016/s0010-0277(02)00004-5

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: The MIT Press.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661–667. https://doi.org/10.1111/j.0956-7976.2004.00737.x

Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, *38*, 1–71. https://doi.org/10.1017/S0140525X14000090

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. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261–1277. https://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. https://doi.org/10.1111/j.1467-8624.2011.01618.x

Koenig, M. A., & Woodward, A. L. (2010). 24-month-olds' sensitivity to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*, 815–826. https://doi.org/10.1037/a0019664

Koenig, M. A., & Woodward, A. L. (2012). Toddlers learn words in a foreign language: The role of native vocabulary knowledge. *Journal of Child Language*, *39*, 322–337. https://doi.org/10.1017/S0305000911000067

Kondrad, R. L., & Jaswal, V. K. (2012). Explaining the errors away: Young children forgive understandable semantic mistakes. *Cognitive Development*, *27*, 126–135. https://doi.org/10.1016/j.cogdev.2011.11.001

Krogh-Jespersen, S., & Echols, C. H. (2012). The influence of speaker reliability on first versus second label learning. *Child Development*, *83*, 581–590. https://doi.org/10.1111/j.1467-8624.2011.01713.x

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674. https://doi.org/10.1037/0033-295X.96.4.674

Kuhn, D., & Dean, D. Jr (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*, 866–870. https://doi.org/10.1111/j.1467-9280.2005.01628.x

Kushnir, T., & Koenig, M. A. (2017). What I don't know won't hurt you: The relation between professed ignorance and later knowledge claims. *Developmental Psychology*, *53*, 826–835. https://doi.org/10.1037/dev0000294

Kuzyk, O., Grossman, S., & Poulin-Dubois, D. (2019). Knowing who knows: Metacognitive and causal learning abilities guide infants' selective social learning. *Developmental Science*, *23*, e12904. https://doi.org/10.1111/desc.12904

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. https://doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development*, *84*, 1253–1268. https://doi.org/10.1111/cdev.12029

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, *126*, 198–212. https://doi.org/10.1016/j.jecp.2014.03.001

Legare, C. H., Sobel, D. M., & Callanan, M. (2017). Causal learning is collaborative: Examining explanation and exploration in social contexts. *Psychonomic Bulletin & Review*, *24*, 1548–1554. https://doi.org/10.3758/s13423-017-1351-3

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*, 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lucas, A. J., Lewis, C., Pala, F. C., Wong, K., & Berridge, D. (2013). Social-cognitive processes in preschoolers' selective trust: Three cultures compared. *Developmental Psychology*, *49*, 579–590. https://doi.org/10.1037/a0029864

Luchkina, E., Morgan, J., Williams, D. J., & Sobel, D. M. (in press). Questions can answer questions about the mechanisms of preschoolers' selective word learning. *Child Development*.

Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science*, *21*, e12663. https://doi.org/10.1111/desc.12663

Luu, B., de Rosnay, M., & Harris, P. L. (2013). Five-year-olds are willing, but 4-year-olds refuse, to trust informants who offer new and unfamiliar labels for parts of the body. *Journal of Experimental Child Psychology*, *116*, 234–246. https://doi.org/10.1016/j.jecp.2013.06.003

MacDonald, K., Schug, M., Chase, E., & Barth, H. (2013). My people, right or wrong? Minimal group membership disrupts preschoolers' selective trust. *Cognitive Development*, *28*, 247–259. https://doi.org/10.1016/j.cogdev.2012.11.001

Mangardich, H., & Sabbagh, M. A. (2018). Children remember words from ignorant speakers but do not attach meaning: evidence from event-related potentials. *Developmental Science*, *21*, e12544. https://doi.org/10.1111/desc.12544

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, *112*, 367–380. https://doi.org/10.1016/j.cognition.2009.05.012

McDonald, K. P., & Ma, L. (2015). Dress nicer = know more? Young children's knowledge attribution and selective learning based on how others dress. *PLoS One*, *10*, e0144424. https://doi.org/10.1371/journal.pone.0144424

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

Nurmsoo, E., & Robinson, E. J. (2009). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science*, *12*(1), 41–47. https://doi.org/10.1111/j.1467-7687.2008.00750.x

Pasquini, E. S., Corriveau, K. H., Koenig, M. A., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perszyk, D. R., Ferguson, B., & Waxman, S. R. (2018). Maturation constrains the effect of exposure in linking language and thought: Evidence from healthy preterm infants. *Developmental Science*, *21*, e12522. https://doi.org/10.1111/desc.12522

Piaget, J. (1930). *The child's concept of the world*. London, UK: Routldge & Kegan Paul.

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology*, *28*, 785–798. https://doi.org/10.1348/026151009X479178

Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development*, *24*(1), 61–69. https://doi.org/10.1016/j.cogdev.2008.07.004

Reid, T. (2000). *An inquiry into the human mind on the principles of common sense*. Edinburgh, UK: Edinburgh University Press. (Original work published 1764)

Robinson, E. J., Fox, A., & Einav, S. (2013). Reading to learn: Prereaders' and early readers' trust in text as a source of knowledge. *Developmental Psychology*, *49*, 505–513. https://doi.org/10.1037/a0029494

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, *85*, 956–970. https://doi.org/10.1111/cdev.12166

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust 1. *Journal of Personality*, *35*, 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*, 443. https://doi.org/10.1037/h0031464

Sabbagh, M. A., & Shafman, D. (2009). How children block learning from ignorant speakers. *Cognition*, *112*, 415–422. https://doi.org/10.1016/j.cognition.2009.06.005

Sabbagh, M. A., Wdowiak, S. D., & Ottaway, J. M. (2003). Do word learners ignore ignorant speakers? *Journal of Child Language*, *30*, 905–924. https://doi.org/10.1017/S0305000903005828

Schult, C. A., & Wellman, H. M. (1997). Explaining human movements and actions: Children's understanding of the limits of psychological explanation. *Cognition*, *62*, 291–324. https://doi.org/10.1016/S0010-0277(96)00786-X

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*, 382–389. https://doi.org/10.1016/j.tics.2012.06.004

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*, 278–290. https://doi.org/10.1016/j.cogdev.2008.01.003

Shultz, T. R., Wells, D., & Sarda, M. (1980). Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology*, *19*, 301–310. https://doi.org/10.1111/j.2044-8260.1980.tb00357.x

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology*, *22*(1), 37–58. https://doi.org/10.1348/026151004772901104

Sobel, D. M., & Austerweil, J. (2016). Coding choices affect performance on a false belief measure. *Cognitive Development*, *40*, 9–23. https://doi.org/10.1016/j.cogdev.2016.08.002

Sobel, D. M., Buchanan, D. W., Butterfield, J., & Jenkins, O. C. (2010). Interactions between causal models, theories, and social cognitive development. *Neural Networks*, *8-9*, 1060–1071.

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development*, *81*, 669–679. https://doi.org/10.1111/j.1467-8624.2009.01422.x

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. https://doi.org/10.1037/a0034191

Sobel, D. M., & Legare, C. H. (2014). Causal learning in children. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 413–427. https://doi.org/10.1002/wcs.1291

Sobel, D. M., & Macris, D. M. (2013). Children's understanding of speaker reliability between lexical and syntactic knowledge. *Developmental Psychology*, *49*, 523–532. https://doi.org/10.1037/a0029658

Sobel, D. M., Sedivy, J., Buchanan, D. W., & Hennessy, R. (2012). Speaker reliability in preschoolers' inferences about the meanings of novel words. *Journal of Child Language*, *39*, 90–104. https://doi.org/10.1017/S0305000911000018

Stephens, E. C., & Koenig, M. A. (2015). Varieties of testimony: Children's selective learning in semantic versus episodic domains. *Cognition*, *137*, 182–188. https://doi.org/10.1016/j.cognition.2015.01.004

Taylor, M. G. (2013). Gender influences on children's selective trust of adult testimony. *Journal of Experimental Child Psychology*, *115*, 672–690. https://doi.org/10.1016/j.jecp.2013.04.003

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*, 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Tong, Y., Wang, F., & Danovitch, J. (2019). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, *23*, e12895. https://doi.org/10.1111/desc.12895

Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants monitor the reliability of potential informants. *Psychological Science*, *25*, 1730–1738. https://doi.org/10.1177/0956797614540178

Turner, C., Giraldeau, L. A., & Flynn, E. (2017). How does the reliability of a model affect children's choice to learn socially or individually? *Evolution and Human Behavior*, *38*, 341–349. https://doi.org/10.1016/j.evolhumbehav.2016.11.005

Van Reet, J., Green, K. F., & Sobel, D. M. (2015). Preschoolers' theory-of-mind knowledge influences whom they trust about others' theories of mind. *Journal of Cognition and Development*, *16*, 471–491. https://doi.org/10.1080/15248372.2014.892875

Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science*, *17*, 443–451. https://doi.org/10.1111/desc.12134

Vanderbilt, K. E., Ochoa, K. D., & Heilbrun, J. (2018). Consider the source: Children link the accuracy of text-based sources to the accuracy of the author. *British Journal of Developmental Psychology*, *36*, 634–651. https://doi.org/10.1111/bjdp.12247

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development: An International Journal of Research and Practice*, *18*(1), 61–71.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*, 270–276. https://doi.org/10.1111/j.1467-7687.2004.00345.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: MIT Press.

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, *25*, 177–182. https://doi.org/10.1177/0963721416645512

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*, 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Woolley, J. D., & Ghossainy, E. M. (2013). Revisiting the fantasy–reality distinction: Children as naïve skeptics. *Child Development*, *84*, 1496–1510. https://doi.org/10.1111/cdev.12081

Yow, W. Q., & Li, X. (2018). The influence of language behavior in social preferences and selective trust of monolingual and bilingual children. *Journal of Experimental Child Psychology*, *166*, 635–651. https://doi.org/10.1016/j.jecp.2017.09.019