CHAPTER ELEVEN

Unifying Pedagogical Reasoning and Epistemic Trust

Baxter S. Eaves Jr.[1] **and Patrick Shafto**

Department of Psychological and Brain Sciences 317 Life Sciences Building, University of Louisville, Louisville, Ky 40292, USA

[1]Corresponding author: E-mail: b0eave01@louisville.edu

Contents

Abstract

Researchers have argued that other people provide not only great opportunities for facilitating children's learning but also great risks. Research on pedagogical reasoning has argued children come prepared to identify and capitalize on others' helpfulness to teach, and this pedagogical reasoning allows children to learn rapidly and robustly. In contrast, research on epistemic trust has focused on how the testimony of others is not constrained to be veridical, and therefore, children must be prepared to identify which informants to trust for information. Although these problems are clearly related, these two literatures have, thus far, existed relatively independently of each other. We present a formal analysis of learning from informants that unifies and fills gaps in each of these literatures. Our analysis explains why teaching—learning from a knowledgeable and helpful informant—supports more robust inferences. We show that our account predicts specific inferences supported in pedagogical situations better than a standard account of learning from teaching. Our analysis also suggests that epistemic trust

Advances in Child Development and Behavior, Volume 43
ISSN 0065-2407,
http://dx.doi.org/10.1016/B978-0-12-397919-3.00011-3

should depend on inferences about others' knowledge and helpfulness. We show that our knowledge and helpfulness account explains children's behavior in epistemic trust tasks better than the standard knowledge-only account. We conclude by discussing implications for development and outline important questions raised by viewing learning from testimony as joint inference over others' knowledge and helpfulness.

One of the most remarkable aspects of human learning is the ability of children to learn so much, so quickly. This ability defies the common wisdom from learning theory, where research has suggested that learning should be impossibly hard (e.g. Gold, 1967). Indeed, humans' ability to learn is so robust that we, but not other animals, are able to accumulate knowledge over generations (Tomasello, 1999). What underlies these remarkable abilities?

One proposed explanation for these impressive feats of learning is an intrinsic understanding of teaching, termed *natural pedagogy* (Csibra & Gergely, 2009). Csibra and Gergely (2009) proposed that people spontaneously engage in, and that children come prepared to identify and understand, acts of teaching. In short, they argue that pedagogy is indicated by ostensive cues—forming joint attention, speaking in child-directed tones, etc.—and in these situations, the information presented is understood to be purposefully communicated and generalizable.

In an effort to understand why pedagogical situations might afford more rapid learning, recent research has presented a formal analysis of pedagogical data selection and its implications for learning, instantiated in a computational model (Shafto & Goodman, 2008). Pedagogical reasoning is formalized as a two-part problem: from the teacher's perspective, which data should be chosen for the learner, and from the learner's perspective, which inferences are afforded by the teacher's choices. The teacher is assumed to be knowledgeable and helpful—she knows the correct hypothesis and chooses examples to increase the learner's belief in that hypothesis. The learner is assumed to know that the teacher is knowledgeable and helpful. The learner then updates her beliefs accordingly. Recent research has investigated the predictions of the model, suggesting that children make stronger inferences from pedagogically chosen data as predicted by the model (Bonawitz et al., 2011; Buchbaum, Griffiths, Gopnik, & Shafto, 2011).

Pedagogical reasoning assumes that informants are trustworthy, but children cannot simply trust everyone they encounter. Recent research on epistemic trust has investigated how children identify which informants to trust for information. Koenig and Harris (2005) showed that by 4 years of age children reliably preferred previously correct informants over incorrect informants in a word-learning task. Subsequent research has shown that children make inferences about informants based on relative accuracy (Fitneva & Dunfield, 2010; Pasquini, Corriveau, Koenig, & Harris, 2007), group consensus (Corriveau, Fusaro, & Harris, 2009), informant familiarity (Corriveau & Harris, 2009), expertise (Sobel & Corriveau, 2010), and more (Fusaro & Harris, 2008; Jaswal & Neely, 2006; Mascaro & Sperber, 2009; Kinzler, Corriveau, & Harris, 2011; Nurmsoo & Robinson, 2009; VanderBorght, 2009).

Children's success on epistemic trust tasks is generally interpreted as reflecting their ability to track informants' knowledge. However, there is reason to believe that knowledge is not the only factor at play. Intuitively, the simple fact that someone is knowledgeable does not preclude them from deceiving. Indeed, a parallel line of research has suggested that 4-year old children are also able to reason about informants' mal-intentions (Mascaro & Sperber, 2009). Specifically, children are able to use behavioral cues such as violence as well as information from other informants—e.g. *that guy is a liar*—to make judgments about informants' reliability. This raises the possibility that 4-year olds' performance in epistemic trust may not be simply attributable to inferences about knowledge alone.

We propose that pedagogical reasoning and epistemic trust are two sides of the same coin. We present a unified framework, within which pedagogical reasoning is a special case of a broader set of models which allow informants to be knowledgeable or not and helpful or not (Shafto, Eaves, Navarro, & Perfors, 2012). We will show how this model can account for learning in pedagogical settings and findings from the literature on epistemic trust, by focusing on specific examples from these literatures. We conclude by discussing implications for cognitive development, connections to related areas of research, and important future directions.

1. A UNIFIED FRAMEWORK OF EPISTEMIC TRUST AND PEDAGOGY

In pedagogical reasoning, informants are assumed to be knowledgeable and helpful; learners use this assumption to guide learning. In epistemic trust, informants may be knowledgeable or not or helpful or not; learners must simultaneously make inferences about the world and about their informants. Therefore, a unified framework must formalize the behavior of different kinds of informants and specify how learners leverage an informant's testimony when the informant's kind is known and when the informant's kind is unknown.

Recent work has formalized aspects of these problems. Shafto and Goodman (2008) proposed a model of pedagogical sampling. Their model formalizes teaching by a knowledgeable and helpful informant as choosing data that tend to maximize the learner's belief in the correct hypothesis and learning as updating one's beliefs assuming that the data have been chosen by a knowledgeable and helpful teacher. Shafto et al. (2012) proposed a model of epistemic trust, where learners simultaneously learn about the world and infer whether informants are knowledgeable or not and helpful or not. Our goal here is to sketch the general framework that unifies these models and to show how this provides a single account for children's behavior across these tasks.

We begin by sketching the modeling framework. We then consider two classes of behavioral tasks that can be captured by the model—pedagogical learning and epistemic trust—and contrast current theoretical accounts with the account offered by the model. By accounting for data across an array of recent work in pedagogy and trust, we unify learning in these scenarios under a common framework.

1.1. The Unified Framework

A model of learning from informants needs to capture two things: how informants select data and how learners learn from different kinds of informants. We adopt a standard probabilistic learning framework (Tenenbaum, Griffiths, & Kemp, 2006). In probabilistic learning, the learner's goal is to update their belief about a hypothesis given data. Bayes' rule dictates that these posterior beliefs are proportional to the product of two quantities: the prior probability of the hypothesis and the probability of observing the data given the hypothesis is true. Thus, given a generative model for the data—a model that specifies how hypotheses are selected and how data are sampled given hypotheses—Bayes' rule specifies how to invert the process—how to infer the hypothesis and the sampling model, given the data.

Generally, models of learning assume that data are *randomly sampled*, that is, data are sampled in proportion to their consistency with the hypothesis. In social learning, it seems that random sampling is rarely applicable. People *choose* data purposefully. Data are not sampled based on their consistency with the hypothesis, but based on the informant's helpfulness, given her knowledge. The key challenge here is to formalize how different kinds of informants produce data. To do this, we must specify how knowledgeability and helpfulness relate to the choices informants make.

Figure 11.1 presents two graphical models depicting how helpfulness and knowledgeability relate to informants' actions in causal and word learning. Graphical models are a powerful tool for defining causal relationships among variables (Pearl, 2009; Spirtes, Glymour, & Scheines, 1993). The fundamental components of graphical models are nodes and edges. A node represents a latent or observed variable; an edge represents a conditional dependency between nodes. Edges are directed and point from parents to children.

Here the goal is to specify how two latent variables corresponding to informants' knowledgeability, *k*, and helpfulness, *h*, affect their choice of action (see Fig. 11.1). Knowledgeability determines the relationship between the informant's beliefs, *b*, and the true state of the world w'. Helpfulness determines the types of actions, *a*, informants choose based on their beliefs. Actions in turn produce effects, *e*, based on the state of the world.

More specifically, knowledgeability, *k*, and helpfulness, *h*, are binary-valued variables corresponding to knowledgeable/naïve and helpful/unhelpful. Beliefs, *b*, model informants' beliefs about the world, *w*; *b* belongs to *B*, the set of possible beliefs; *w* belongs to *W*, the set of states of the world. In word learning, *B* and *W* are sets of labels, and in causal learning, they are sets of causal graphs. For example, in a game in which an informant points to one of two cups under which a ball is hidden, *B* and *W* would both be composed of the set of possible locations of the ball, {*cup*1, *cup*2}. This task

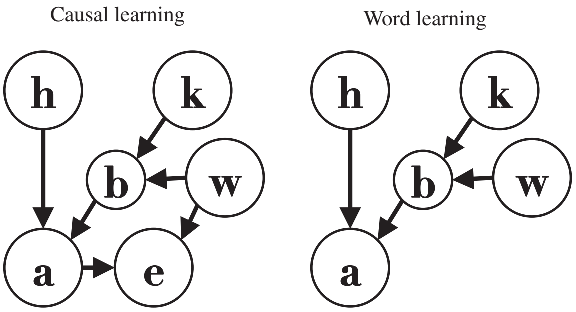

Figure 11.1 Graphical representation of the model. On the left is the causal learning model. On the right is the word-learning model. In causal learning, an informant's action, *a*, is an intervention on the world, *w*, which elicits a response from the world: an effect, *e*. In word learning, actions are labels and do not affect the world; thus, the effect node is not present in the word learning model.

can be thought of as a labeling task because the informant labels a cup *as the one containing the ball.*

Knowledgeability specifies the relationship between the informant's beliefs and the world. If an informant is knowledgeable, then her beliefs, b, correspond to the true state of the world, w'; she would know which cup the ball is under. In contrast, if an informant is naive, then her beliefs are uniform over the set of possible beliefs; she would not know which cup the ball was under. More formally,

$$P_I(b = w'|k) = \begin{cases} 1 & \text{if knowledgeable} \\ 1/|W| & \text{if naïve} \end{cases}, \qquad (11.1)$$

where $|W|$ is number of possible hypotheses about the world.

Actions, a, are chosen based on the informant's helpfulness and beliefs. If an informant is helpful, she will act to maximize the learner's belief in the belief she holds; if she is not helpful, she will act to minimize the learner's belief in the belief she holds. Formally,

$$P_I(d|b_I) \propto P_L(b_L = b_I|d)^{\alpha}, \qquad (11.2)$$

where

$$\alpha = \begin{cases} 1 & \text{if helpful} \\ -1 & \text{if not helpful.} \end{cases} \qquad (11.3)$$

When α is 1, the informant chooses data that tend to lead the learner to her belief. When α is -1, the informant chooses data that tend to lead the learner away from her belief. In the cup game, a knowledgeable and helpful informant would point to the cup she believed the ball was hidden under; a knowledgeable but unhelpful informant would point to the cup opposite the one she believed the ball was hidden under. Because actions are based on informants' beliefs, and beliefs are based on informants' knowledgeability, naive informants, regardless of helpfulness, will appear to produce actions scattershot. In the cup game, a naive informant will point to the correct cup half of the time. The helpful naive informant points at the correct cup because she has guessed correctly and the unhelpful naive informant points at the correct cup because she believes the ball is under the wrong cup and attempts to lead the learner away from it.

In causal learning (see Fig. 11.1), there is an additional factor, the effects of actions. The effects of actions are determined by the action chosen and the true causal structure of the world. In word learning (see Fig. 11.1), the action is an utterance, and because an utterance does not affect the world, the effect node is removed.

This sampling model allows us to consider which actions are likely to be chosen by different kinds of informants and, given actions, allows learners to infer what kind of informant they are dealing with. In different social (and experimental) scenarios, informants' helpfulness and knowledgeability may implicitly take on certain values. When the informant claims to be knowledgeable and helpful or if certain social cues are present (Csibra & Gergely, 2009), learners may assume the informant is knowledgeable and helpful. Similarly, learners may be exposed to cues which lead them to believe an informant is knowledgeable and unhelpful (deceptive). Importantly, Bayesian inference allows us to learn who to trust and what to infer from informants' actions.

1.2. Modeling Pedagogical Learning

To illustrate how the model accounts for learning from pedagogically selected data, we consider two sets of results. The first is from Shafto and Goodman (2008), which examined pedagogical sampling and how it affects what is learned, and the second is from Bonawitz et al. (2011), which looked at how pedagogy affects future exploration. In both these cases, we contrast the model predictions with the strong sampling proposal offered by Xu and Tenenbaum (2007).

In Shafto and Goodman (2008), participants played a concept learning game in which they were to locate a hidden rectangle using pairs of points labeled as inside (positive) or outside the rectangle (negative). In the pedagogical sampling condition, participants both taught and learned. When teaching, participants observed the rectangle and chose points to mark. When learning, participants saw labeled points on a blank screen and then drew a rectangle which they believed was the actual rectangle the teacher intended. In the non-pedagogical sampling condition, participants searched for the rectangle themselves. They observed a blank screen and chose points to have labeled. Once the points were placed on the screen, the software labeled them as in or out of the rectangle.

The results showed that in the pedagogical conditions teachers chose to place pairs of positive examples at opposite corners and pairs of negative examples at opposite edges or corners. The intuition here is that teachers choose points to maximize a learner's belief in a single hypothesis and placing the positive examples at the opposite corners rules out all rectangles smaller than the actual rectangle (see also Rhodes, Gelman, & Brickman, 2010).

Learners' inferences in the pedagogical condition showed a distinct pattern. Rectangles were drawn with the two positive examples in the corners and with edges close to negative examples. In the nonpedagogical sampling condition, rectangles drawn by participants did not show any discernible pattern, suggesting that learners in the pedagogical condition inferred that teachers choose data purposefully—positive examples in the corners of the rectangle, and negative examples at the boundaries—while learners in the non-pedagogical condition did not.

Previous accounts of word learning have suggested that learners' inferences from teachers' demonstrations in word learning may be modeled by assuming strong sampling (Xu & Tenenbaum, 2007; see also Tenenbaum, 1999). In strong sampling, learners assume that examples are selected randomly *from the true concept*. Because this creates a natural preference for smaller concepts (that are consistent with the examples), given only positive examples, learners will rapidly converge to the correct concept. Strong sampling can, therefore, account for learning from positive examples. But because examples are assumed to be randomly sampled, it cannot explain teachers' preference for examples in the corners. Similarly, because strong sampling assumes that only positive examples are chosen, it cannot explain negative example selection or learning from negative examples.

Under our model, a teacher is someone who is knowledgeable, $k = 1$, and helpful, $h = 1$; they know about the world, and they want learners to as well. Teachers therefore choose data to increase learners' beliefs in the true state of the world (see Eqn 11.2). In this case, because the teacher is knowledgeable, her belief is assumed to match the true state of the world, $b_i = w'$, and because she is helpful, the exponent, α, is 1. In the rectangle game, W and B are both the set of possible rectangles, where b is a single rectangle, and w' is the true rectangle. Possible data are the set of possible pairs of negative or positive examples. For positive examples, if a teacher chooses narrow data, positive examples closer to the center of the true rectangle, they rule out fewer incorrect rectangles than if they choose data in the corners, $P(w'|D = \text{narrow}) < P(w'|D = \text{corners})$, and therefore, $P(D = \text{narrow}|w') < P(D = \text{corners}|w')$.

Negative examples should be chosen to constrain the number of possible rectangles. Choosing negative examples at the sides rules out all rectangles larger than the boundaries of the points. As before, placing negative examples away from the rectangle boundaries (wide data) rules out fewer rectangles larger than the target and makes learning the target rectangle less likely, $P(w'|D = \text{wide}) < P(w'|D = \text{sides})$, and therefore, $P(D = \text{wide}|w') < P(D = \text{sides}|w')$. Because learners use their knowledge of how points are chosen, when points are chosen at random (nonpedagogical condition), the model cannot make assumptions about *why* specific points were chosen and therefore chooses randomly based on the examples present.

The model explains why faster learning is achieved in pedagogical learning. Using their knowledge of how teachers choose data, people are able to infer the correct rectangle with two points, rather than six perfectly placed points (four negative examples, one at each side to constrain the maximum size and two positive examples at opposite corners to constrain the minimum size). One implication of this increased confidence is that after observing pedagogically sampled data, one may be less curious than after observing the same data chosen in a nonpedagogical setting. Bonawitz et al. (2011) explored this possibility: would learners presented with pedagogically sampled data be less likely to search for additional data?

Children were presented with a novel, complex-looking toy. Unbeknownst to the children, the toy was built to have four nonobvious functions: a knob that caused squeaking, a key that made music, a button that turned on a light, and a tube with a mirror that reversed the child's face. The toy was designed to appear complex looking to lead children to believe that there could be many functions of the toy.

Children were randomly assigned to one of a number of conditions. We focus on two: the pedagogical condition and the accidental condition. These conditions were set up such that children observed the same data: pulling a knob causes squeaking. Across conditions, the social context was manipulated. In the pedagogical condition, the demonstrator was presented as knowledgeable (stating, "This is my toy") and helpful [via pedagogical cues such as establishing joint attention, repeating the child's name, etc. (Csibra & Gergely, 2009)]. In the accidental condition, the demonstrator was presented as naive (saying, "Look at this toy I found") and the demonstration was presented as accidental. As the demonstrator put the toy down, their hand hit the knob, causing a squeak. In both conditions, there were two demonstrations to ensure that the child saw the cause of the squeak. After the demonstration, the child was allowed to play with the toy. Experimenters tracked various measures of how much exploration children engaged in as well as the total number of built-in functions children discovered. The results showed that children in the pedagogical condition explored less and discovered fewer functions of the toy than did children in the accidental condition.

Under strong sampling, data are selected randomly from the true concept. However, strong sampling does not specify how much data to select. Therefore, strong sampling offers no explanation as to why a specific number of demonstrations are better than any other.

To explain these results under our model, we must specify the possible beliefs/states of the world and data. Possible states of the world (and beliefs) include different numbers of possible functions. W includes the possibility that the toy has no functions $w = 0$, one function $w = 1$, two functions $w = 2$, etc. Possible actions include no demonstration $a = 0$, performing one action $a = 1$, two actions $a = 2$, etc. In the experiment, the question is what should a learner infer from the teacher's choice to *only* demonstrate that pulling the knob leads to squeaking. Intuitively, given the teacher is knowledgeable and helpful, if the toy had any other functions, we would expect the teacher to have shown them to us. The model predicts that, given a toy with n functions, i.e. $w' = n$, we would expect n demonstrations, $a = n$. Consider what would happen if the teacher demonstrated only $n - 1$ functions. The learner could rule out all hypotheses in which there are less than $n - 1$ functions. However, all hypotheses with n or more functions are still possible. By demonstrating one more function, the teacher would eliminate one more possibility, increasing the learner's belief in n functions, $P(w = n|a = n) > P(w = n|a = n - 1)$. Thus, the model predicts that teachers demonstrate all functions, $P(a = n|w' = n) > P(a = n - 1|w' = n)$, and given such a demonstration, learners infer no more functions exist, $P(w = n|a = n) > P(w = n + 1|a = n)$.[1] Because the chance that other functions exist is low, there is no need to spend time looking for them.

On the other hand, when the demonstrator accidentally elicits a squeaks, the data rule out the possibility that there are zero functions, but because the action and effect were a result of a chance occurrence (random sampling), one cannot assume there are not more functions. In this case, if one wishes to learn about the toy, one must explore.

[1] This discussion assumes that all hypotheses are equally likely. The assumption is made for expository simplicity, and the conclusions hold across a range of scenarios.

1.3. Modeling Epistemic Trust

The previous section focuses on situations in which the knowledge and intent of the informant are known (or can be reasonably assumed). Of course, that is not the problem that people typically face in the world. Informants may or may not be trustworthy, and research shows that children track who to trust. We present a standard account of children's reasoning—the knowledge heuristics account—and contrast it with our own. We begin by discussing three representative findings from epistemic trust literature. Finally, we will show how our model of learning from informants can account for all the results discussed and, thus, all corresponding heuristics.

The trust tasks examined in this section each follow a similar format. Learners are given some demonstrations, which they can use to make inferences about their informants. For example, in Pasquini et al. (2007), informants label four common objects with varying accuracy; in Corriveau, Fusaro, et al. (2009), several informants point to an object after hearing a label given by an experimenter. After the demonstration, learners must choose which informant to ask or which informant's information to endorse when faced with a novel object or label (novel trial). The key question is whether children show systematic preferences for different informants, and if so, what kinds of experience lead children to choose one informant over another?

In Pasquini et al. (2007), kids observed two informants label four common objects, such as a ball or a shoe, with varying accuracy: 100%, 75%, 25%, or 0%. After these familiar trials, children were presented a novel object. In *ask* trials, children asked one of the two informants for the label, and in *endorse* trials, both informants labeled the object with different labels, and the child was then asked which she thought the object was called. The results showed that children indeed form preferences for more accurate informants, meaning children prefer to ask, or endorse the label given by more accurate informants more often. Qualitatively, for both 3- and 4-year olds, the results showed that the preference for the more accurate informant decreased with the relative accuracy of the informants, e.g. children in the 75% versus 0% accurate condition showed a higher preference for the 75% accurate informant than did children in the 75% versus 25% accurate condition. However, whereas 3-year olds showed less and less differentiation across the 100% versus 0%, 100% versus 25%, 75% versus 0%, and 75% versus 25% conditions, 4-year olds at minimum show a sharp differentiation of the 75 versus 25 condition from the others and appear to have somewhat improved performance in the 100% versus 0% condition relative to the others.

To explain these differences, Pasquini et al. (2007) suggested that children's choices are guided by heuristic monitoring of the inaccuracy of the informant. That is, to begin both informants are categorized as trustworthy, but when an informant labels incorrectly, that informant is categorized as inaccurate. According to Pasquini et al. (2007), for 3-year olds, the strategy stops here. An informant is either accurate or inaccurate, and this binary explanation accounts for 3-year olds' poor performance when both informants have labeled one or more object inaccurately. To explain the differences between 3- and 4-year olds, Pasquini et al. (2007) propose that 4-year olds also use the frequency of informants' mislabelings and are thus better able to choose between inaccurate informants than 3-year olds.

Corriveau and Harris (2009) carried out an experiment nearly identical to Pasquini et al. (2007), with two differences: one of the informants was the child's preschool teacher, and rather than parametrically altering accuracy on familiar object trials, informants labeled 100% or 0% accurately. When the child first encountered the two informants, one novel and one familiar, the child was presented with a novel object and answered ask and endorse questions. Both 3- and 4-year-old children preferred the familiar informant. After the novel trials, children observed familiar object trials in which the familiar informant labeled 100% accurately and the novel informant labeled 0% accurately or the familiar informant labeled 0% accurately and the novel informant labeled 100% accurately. Four-year olds preferred the familiar informant after having seen her label correctly, more so than in novel trials. When the familiar informant labeled incorrectly, 4-year olds preferred the novel informant who had labeled correctly. Three-year olds still preferred the familiar informant even when she had labeled incorrectly.

According to the heuristic account proposed by Pasquini et al. (2007), all informants initially belong to the trustworthy category. If this were true, given a novel and a familiar informant, children should choose both informants equally because they are both trustworthy. Corriveau and Harris (2009) suggest that perhaps children have witnessed the familiar informant label accurately many times in the past and have some bias toward accurate information which would create the familiarity bias. Under this proposal, children must be tracking some kind of frequency of correct answers. Given that 3-year olds also show a preference for familiar informants, this creates a contradiction with the previous experiment, where their behavior was explained by not attending to the frequency information but by categorization. To explain the current results, it seems necessary to propose that familiarity is an additional heuristic that guides children's choices in the pretest.

For the posttest (once the informants have labeled familiar objects), a familiarity and accuracy account would have to specify how these two factors interact. When the familiar informant labeled accurately and the novel informant labeled inaccurately, 4-year olds' preference for the familiar informant increased compared to the pretest, but 3-year olds' preference remained the same. When the familiar informant labeled inaccurately and the novel informant labeled accurately, 4-year olds then preferred the novel informant, but 3-year olds continued to prefer the familiar informant. If children used the raw frequency of informant's truthful productions, we would expect to see preferences for the familiar informant to remain for both groups even when the familiar informant labeled inaccurately, as it would be reasonable to assume that a teacher has produced enough truthful information (likely hundreds of productions) to outweigh four mislabelings. Accordingly, Corriveau and Harris (2009) suggest that there is a bias for recent accuracy as well. This, however, does not explain 3-year olds' continued preference for the inaccurately labeling familiar informant or why their preference for the familiar informant does not increase when she labeled accurately. For this reason, the authors suggest that for 3-year olds, familiarity, and not the productions of information that comprise it, outweighs accuracy as a heuristic.

Corriveau, Fusaro, et al. (2009) looked at how children choose informants and data when learning about novel objects from a group of novel informants given only a set of novel labels. Four informants are presented with three novel objects. An experimenter asks, "Show me the modi" after which, each informant points to an object. Three informants agree and one dissents. This occurs for several trials. On each trial, the same informants agree, and the same informant dissents. After the informants have pointed, the child is asked which she believes is the modi. Here, learners have only labels from a few informants by which to make inferences and therefore cannot use an inaccuracy strategy. The results showed that children prefer the object indicated by the majority and that there were no differences in age groups. After group trials, children participated in novel object labeling trials in which one informant was from the majority and the other was the dissenter. Again, children preferred the informant from the majority, and there was no effect of age.

Corriveau, Fusaro, et al. (2009) argue that children prefer informants who are part of a broader consensus and that children may believe informants from a majority are more epistemically trustworthy or may otherwise form some kind of emotional attraction to non-dissenters. In other words, children exhibit a heuristic majority bias: when learners have only a set of novel object labels, they choose the one that is most agreed upon. Note that this bias cannot be derived from previous biases. The accuracy bias cannot be applied because there is only novel information, so learners cannot judge the accuracy of the information; the familiarity bias cannot be applied because all the informants are novel. Also note that in this study, no developmental differences were observed, and therefore, no developmental change in this ability was proposed.

These studies paint an interesting picture of children's abilities: they show remarkable subtlety in reasoning, with developmental differences in some cases, but not others. For each subtle variation in behavior, the heuristic account proposes more heuristics, leading to complex and often underspecified interactions given a specific scenario or a developmental stage. The accuracy bias works differently for 3- and 4-year olds. When at least one informant is familiar, it works differently still for 4-year olds and not at all for 3-year olds. When groups of informants are involved, 3- and 4-year olds do not differ, they use the same heuristic of choosing with the majority. Similarly, it is not clear how the existing heuristics apply in minimally different scenarios. If a dissenting informant were familiar, which heuristic would children use: majority or familiarity? Would this change with age? What is needed is an account that provides a more parsimonious explanation of existing phenomena and makes principled generalizations across scenarios.

We propose that behavior can be understood as joint inference about informants' knowledge and intent (Shafto et al., 2012). The model observes informants' actions and decides which kind of informant is most likely to have produced those actions, e.g. helpful/unhelpful, knowledgeable/naive. On novel trials, the model uses what it knows about how different types of informants choose data, along with the inferences it has made about its informants, to predict which informant is most likely to produce correct labels in the future.

The model both learns about informants and predicts their future behavior. In Pasquini et al. (2007) and Corriveau and Harris (2009), during familiar object labelings, the values of w' and a are fixed because the objects are familiar and the labels are observed. Learners leverage this information in order to infer k and h, whether an informant is knowledgeable and helpful. Informants who label more accurately are more likely to be helpful. Informants who always label accurately are likely helpful and knowledgeable, and informants who never label correctly are likely knowledgeable, but unhelpful because naive informants, whether helpful or not, will occasionally produce the correct labels. Like in the account proposed by Pasquini et al. (2007), learners use accuracy/inaccuracy to choose informants. However, rather than an ad hoc approach based on tallying correct answers directly, we propose that children are actually inferring unobserved causal properties of informants—whether the informant is knowledgeable and whether the informant is helpful. Developmental differences are explained in our framework as changing assumptions about people. While 4-year olds' behavior is best explained by a model that infers knowledgeability and helpfulness, 3-year olds' behavior is best explained by a model that infers knowledgeability but assumes helpfulness. Based on knowledgeability alone, informants who have mislabeled one or more times become similar. Under the model, this accounts for 3-year olds' performance in choosing between inaccurate informants.

Inferences are made similarly when learning from familiar informants. In the case of Corriveau and Harris (2009), where the informant is a preschool teacher, familiarity is modeled as positive past experience. In contrast with Corriveau and Harris (2009), in our model, this experience manifests as strengthened prior beliefs on k and h (see Equation 11.5 and 11.7) rather than a heuristic assumption of a truth bias. In familiar object labeling trials, learners' preferences are not affected as much by the familiar informant's labels as they are by the novel informant's labels. Learners already have strong beliefs about the familiar informant. Stronger beliefs are more difficult to override; it takes more evidence to do so. Age differences are explained as earlier. Without the ability to account for the helpfulness of informants, the knowledge-only model does not differentiate as much between always and never accurate informants.

The models account for the result in both phases of Corriveau, Fusaro, et al. (2009), and both show similar predictions. Because there are only informants' labels from which to infer the correct label, actions are fixed, and the informants' knowledgeability and helpfulness as well as the correct object must be learned. Because the probability of naive, or not helpful informants converging on the same label is low, the model infers that the agreeing informants are likely knowledgeable and helpful and indicate the correct label. After the model has made inferences about informants' knowledgeability and helpfulness, it can use this information to decide which informants are more likely to label correctly in the future. The model chooses informants based on the probability they will label correctly in the future, accounting for the preference for non-dissenting informants.

Therefore, the majority bias proposed by Corriveau, Fusaro, et al. (2009) is a manifestation of the non-dissenting informant's past labelings, which in the group trials were inferred to be accurate.

Three findings from the epistemic trust literature—parametrically varying preference for accurate informants, variations in preference based of familiarity and accuracy, and preference for informants from groups over dissenters—illustrate differences between a standard heuristic-based account and our modeling framework. Whereas the heuristic account incurs a proliferation of explanations to account for variations depending on the task and children's location on a developmental trajectory, we propose an inference framework that explains variations in children's behavior across tasks in terms of reasoning about informants' knowledge and helpfulness. We showed that the model explains why these heuristics work, and as such, they need not be thought of as heuristics, but as similar inferences under a common mechanism. Children learn about informants underlying epistemic qualities and in turn use what they have learned to infer informants' future accuracy.

2. CONNECTIONS, IMPLICATIONS, AND FUTURE DIRECTIONS

Researchers almost universally agree that other people play a key role in explaining the power of human learning. Researchers also agree that learning from others leaves us potentially vulnerable to misinformation. These two lines of research—on pedagogical reasoning and epistemic trust—have advanced largely independently of each other. We have presented a unified approach in which pedagogical reasoning and epistemic trust are different facets of the same problem: reasoning about other people's knowledge and intent. We have illustrated how our framework predicts pedagogical data selection and its implications for learning and explains children's behavior when learning who to trust for information.

We have contrasted our approach with an account from each of these literatures: strong sampling for pedagogical learning and heuristic monitoring for epistemic trust. In each case, we argue that our model represents an improvement over these previous accounts. Unlike strong sampling, our approach to pedagogical reasoning explains teachers' choices of evidence, learning from negative evidence and learning from variable amounts of data. Unlike the heuristic account, our approach explains variation in children's behavior across situations and developmental stages in terms of a simple set of principles based on reasoning about informants' knowledge and helpfulness.

Together, these arguments illustrate how pedagogical learning and epistemic trust can be viewed as two sides of the same coin. In pedagogical learning, the informant is known to be knowledgeable and helpful, and the goal is to learn about the world. In epistemic trust, often the world is known, and the goal is to learn about the informant. In the former case, knowledge about the informant provides leverage for learning about the world. In the latter, knowledge about the world provides leverage for learning about the informant.

However, as demonstrated by Corriveau, Fusaro, et al. (2009), learning about the world and informants may also occur simultaneously, and our model captures this ability too. This highlights a remarkable ability of children—the ability to perform joint inference (or learning) over multiple variables. While in some ways this appears remarkably sophisticated, this ability is the crux of the explanation for how social learning affects learning about the world; arguably, the problem of childhood is one of learning about both the physical and social worlds.

In the remainder of the paper, we briefly consider connections to previous research, implications for other literatures, and outline potential future directions.

2.1. Connections

Our unified framework suggests that children reason about other people's knowledge and helpfulness. This proposal contrasts with standard work on theory of mind (ToM), where children have been shown to have difficulty reasoning about other people's knowledge (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983; Wellman, Cross, & Watson, 2001). In standard ToM tasks, children must reason about other people's behavior when the person's beliefs are false. In these tasks, results suggest that 3-year-old children have difficulty predicting people's behavior, while 4-year old children do not (but see Onishi & Baillargeon, 2005). The key element of these tasks is that the actor's beliefs are not in accord with the truth while the child's are.

In contrast, in the pedagogical reasoning tasks we considered, the learner does not know the true state of the world and tries to infer it based on the assumption the informant is knowledgeable and known to be helpful, as in pedagogy. Similarly, in epistemic trust tasks, the learner either knows the state of the world and assesses the informant's behavior against that or the learner does not know the true state of the world and assesses multiple informants against each other. In either case, children do not need to predict an informant's behavior based on that informant's false beliefs; informants either have true beliefs or are uncertain. Thus, there is no necessary reason why pedagogical or epistemic trust reasoning necessarily depends on false belief reasoning.

Our approach also differs from previous research modeling aspects of ToM. Butterfield, Jenkins, Sobel, and Schwertfeger (2009) formalized certain aspects of ToM using Markov random fields, providing qualitative arguments that the model can capture effects of uncertainty and reliability and gaze following abilities. Baker, Saxe, and Tenenbaum (2009) formalize action understanding as inverse planning and provide evidence based on adults' judgments about the goals of animated agents in sprite worlds. Unlike Butterfield et al. (2009), our approach has been to not only demonstrate capabilities of our models but to leverage models to provide explanations for developmental changes in performance. Unlike Baker et al. (2009), our focus is on learning about the world and others through intentional acts of communication, as opposed to simple observation.

2.2. Implications

The unified framework covers a wide variety of research and therefore potentially has broad implications. Here, we focus on the two literatures for which it has the most obvious implications: broader literature on epistemic trust and research on deception.

We have focused on the subset of epistemic trust literature which investigates what informant characteristics children track by manipulating the data informants produce. There is an extensive literature suggesting that these are not the only characteristics that children attend to. Children also attend to perceptual aspects of the stimuli (Corriveau, Harris, et al., 2009), informants' accents (Kinzler et al., 2011), and others' nonverbal cues such as bystander reactions (Fusaro & Harris, 2008). Each of these situations leverages additionally information that does not simply reduce to reasoning about the evidence that people provide. Consequently, to model these scenarios would require additional machinery. For instance, with a model of the relationship between perceptual similarity and categories (e.g. Anderson, 1991), the framework could be extended to generate predictions regarding how perceptual similarity of stimuli interacts with judgments about trust. With a model of the relationship between social groups and accents and a distinction between different groups of informants, the framework could be extended to explain the effects of accent on trust. These suggest interesting directions for future research.

There is also a vast body of work which examines children's ability understand and engage in deception. These works cover white lies (Lee & Talwar, 2002), concealing transgressions (Lewis, Stanger, & Sullivan, 1989; Talwar & Lee, 2002), deception games (Chandler, Fritz, & Hala, 1989; Hala, Chandler, & Fritz, 1991; Couillard & Woodward, 1999; Sodian & Frith, 1992), and deception for self-gain (Peskin, 1992). Our model formalizes two types of intentions that a communicative agent may have: helpfulness, and what we have called unhelpfulness. Note that we formalized unhelpfulness as minimizing the learners' belief in the correct hypothesis. This represents a weak case of what may be considered deception—the goal is to mislead the learner.

It is interesting to ask whether the modeling framework may be used to model development of reasoning about deception. A key issue would be identifying cases which have properties similar to the studies we focused on when modeling trust: a simple manipulation of helpfulness and knowledgeability. Couillard and Woodward (1999) designed an experiment in which the informant's helpfulness was left unknown, but could be learned from data, which is a similar design to the epistemic trust studies, where aspects of the informant must be inferred based on the data that they choose. Mascaro and Sperber (2009) follow a format similar to Couillard and Woodward (1999). Here children were told beforehand by the experimenter, in the liar condition, that the informant was a "big liar" and always told lies. Clearly, these are cases where our framework could be applied and used to generate predictions. In the former case, the model would reason about a knowledgeable informant and infer their intent based on the outcome of the trials. In the latter case, the model would make predictions about the outcomes of the trials given the informant's knowledge and intent (Shafto et al., 2012). These examples indicate that systematic investigation of predictions about the development of reasoning about deception is an important direction for future work.

2.3. Future Directions

The literatures on pedagogical reasoning and epistemic trust stand in contrast with each other. The literature on pedagogical reasoning seeks to explain how children could learn so much, so quickly. In order to explain these abilities, Csibra and Gergely (2009) and colleagues (see also Tomasello, 1999; Tomasello, Carpenter, Call, Tanya, & Moll, 2005) have suggested that infants come prepared to identify and interpret acts of teaching. In contrast, the literature on epistemic trust notes that not all informants should be trusted and seeks to explain how children determine who is trustworthy. Thus, while the pedagogy literature emphasizes the need to assume that informants are knowledgeable and helpful, the epistemic trust literature emphasizes the need to assume that informants are *not* always knowledgeable and helpful.

Confounded with this difference in emphasis is a difference in the ages of the children studied. The literature on pedagogy seeks to study children as young as possible [often from 1 year of age and on through school ages (Gergely, Egyed, & Király, 2007; Topál, Gergely, Miklósi, Erdohegyi, & Csibra, 2008)], while the literature on epistemic trust tends to study children 3 years old and up. The differences between these literatures belie the common developmental questions: what assumptions/abilities are built in and what is the developmental trajectory of learning from informants.

There are two main possibilities for resolving these differences. First, it could be that children innately assume informants are knowledgeable and helpful, and this is gradually unlearned through experience with older siblings and tricky grandfathers. Or second, it could be that children begin with weak assumptions about the nature of informants, and their early pedagogical reasoning and later skepticism are both a consequence of their changing experiences with informants and beliefs about the world.

A key question for future research is to characterize and test the consequences of each position, a task that computational modeling is uniquely positioned to facilitate. Our recent research suggests that developmental changes between 3 and 4 years of age on epistemic trust tasks may be attributable to changes in expectations about informants (Shafto et al., 2012). Similarly, computational simulations can be used to ask to what degree can each hypothesis explain the speed of learning and what kinds of developmental trajectory could we expect from each hypothesis? These represent important directions for future research and ways in which computational models and empirical research may mutually inform each other.

3. CONCLUSION

We have presented a unified account of reasoning about learning from pedagogically sampled data and epistemic trust. We propose that these are instances of the broader problem of reasoning about informants' knowledgeability and intent. We illustrated the workings of our framework on representative problems from each literature and contrasted the account provided by our model with theoretical explanations specific to each domain. We suggest that our approach to modeling children's learning and development points to fruitful avenues for future research. There is much to be learned about how other people affect children's learning and development, but we are confident that continued integration of computational modeling and empirical methods points a way forward.

APPENDIX: MODEL SPECIFICATION

Here we describe in detail how the individual components of the model function and interact. We then describe mathematically how the model chooses informants.

Helpfulness and Knowledgeability

Learners' beliefs about helpfulness and knowledgeability can be broken down into three levels: beliefs about informants in general, beliefs about an individual informant, and beliefs about an informant on a given trial. Working from the bottom up, in the model, the informant is knowledgeable on a given trial with probability θ_k or $P(k) = \theta_k$. That is,

$$k \sim \text{Bernoulli}(\theta_k), \tag{11.4}$$

where k describes an informant's knowledgeability on a particular trial and θ_k describes the tendencies of an individual informant.

These tendencies are derived from the learner's prior beliefs about informants in general, which follow a Beta distribution with two hyperparameters: uniformity, $\gamma_k \in (0, \infty)$, and bias, $\beta_k \in (0, 1)$. Uniformity corresponds to the beliefs that people are uniform in their knowledgeability (high uniformity, $\gamma_k \rightarrow \infty$) or that people tend to have different levels of knowledgeability (low value, $\gamma_k \rightarrow 0$). Bias corresponds to the belief that people are knowledgeable $(\beta_k \rightarrow 1)$ or not $(\beta_k \rightarrow 0)$. Putting these pieces together,

$$\theta_k \sim \text{Beta}(\gamma_k \beta_k, \gamma_k(1 - \beta_k)). \tag{11.5}$$

Helpfulness is defined similarly to knowledgeability,

$$h \sim \text{Bernoulli}(\theta_h) \tag{11.6}$$

$$\theta_h \sim \text{Beta}(\gamma_h \beta_h, \gamma_h (1 - \beta_h)). \tag{11.7}$$

State of the World

The true state of the word is distributed uniform over possible states,

$$P(w') = \frac{1}{|W|}, \tag{11.8}$$

where $|W|$ is the number of possible states of the world.

Beliefs

Informants' beliefs are determined by their knowledgeability and the true state of the world. Informants who are knowledgeable have beliefs corresponding to the true state of the world; naive informants have beliefs distributed uniformly over all possible states of the world. Formally,

$$P_I(b = w'|k) = \begin{cases} 1 & \text{if } k \\ 1/|W| & \text{if naïve.} \end{cases} \tag{11.9}$$

Actions

The action performed by an informant is dependent on that informant's beliefs and helpfulness. Here we must specify the model for two types of actions: intervention on a causal device (e.g. Fig. 11.1, left) and labeling (e.g. Fig. 11.1, right). In the case of labeling, a helpful informant will utter the label corresponding to her beliefs; an unhelpful informant will choose any label other that the one corresponding to her beliefs. Formally,

$$P(l|b,h) = \begin{cases} 1 & \text{if } l = b \text{ and } h = 1 \\ 0 & \text{if } l \neq b \text{ and } h = 1 \\ 0 & \text{if } l = b \text{ and } h = 0 \\ 1/(|W| - 1) & \text{if } l \neq b \text{ and } h = 0 \end{cases}. \tag{11.10}$$

In the case of interventions on a causal device, actions are chosen according to Eqn 11.2. Effects are then determined by the intervention and the underlying causal structure of the world.

Learning about and Choosing Informants

The studies in section require learners to choose informants for information (ask trials). As in the studies, we focus here on world learning. The model chooses informants with probability proportionate to how likely they are to produce correct labels in the future given their knowledgeability, helpfulness, and previous experience, E. To do this, the model must predict the probability of each informant labeling correctly for each possible true state of the world, $w\prime\in W$. For each informant,

$$P(l = w'|k, h, E) = \sum_{w'\in W} P(w') \int P(l = w'|w', \theta)P(\theta|\gamma, \beta, E)\mathrm{d}\theta, \tag{11.11}$$

where, for purposes of brevity, $\theta = \theta_h, \theta_k, \gamma = \gamma_h, \gamma_k$, and $\beta = \beta_h, \beta_k$. The integral over θ is not analytically solvable. We therefore approximate using Monte Carlo methods (here, rejection sampling).

The probability in Eqn 11.11 is then normalized over informants. For example, given two informants a and b, the model chooses to ask informant a with probability equal to

$$P(a) = \frac{P_a(l = w'|k, h, d)}{P_a(l = w'|k, h, d) + P_b(l = w'|k, h, d)}. \tag{11.12}$$

Results for endorse trials can be similarly captured by taking inferences summed over informants and normalized over each true state over the world.

REFERENCES

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review, 98*(3), 409.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*, 329–349.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46.

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*(3), 322–330.

Buchbaum, D., Griffiths, T., Gopnik, A., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120*, 331–340.

Butterfield, J., Jenkins, O. C., Sobel, D. M., & Schwertfeger, J. (2009). Modeling aspects of theory of mind with Markov random fields. *International Journal of Social Robotics, 1*, 41–51.

Chandler, M., Fritz, A., & Hala, S. (1989). Small-scale deceit: deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*(6), 1263–1277.

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: weighing familiarity and recent accuracy. *Developmental Science, 12*(3), 426–437.

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: preschoolers prefer nondissenters as informants. *Psychological Science, 20*(3), 372–377.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., et al. (2009). Young children's trust in their mother's claims: longitudinal links with attachment security in infancy. *Child Development, 80*(3), 750–761.

Couillard, N., & Woodward, A. (1999). Children's comprehension of deceptive points. *British Journal of Developmental Psychology, 17*(4), 515–521.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

Fitneva, S. A., & Dunfield, K. A. (2010). Selective information seeking after a single encounter. *Developmental Psychology, 46*(5), 1380–1384.

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science, 11*(5), 771–777.

Gergely, G., Egyed, K., & Kiraly, I. (2007). On pedagogy. *Developmental Science, 10*(1), 139–146.

Gold, E. (1967). Language identification in the limit. *Information and Control, 10*, 447–474.

Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: deception as a marker of three-year-olds' understanding of false belief. *Child Development, 62*(1), 83.

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*(9), 757–758.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106–111.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*(6), 1261–1277.

Lee, K., & Talwar, V. (2002). Emergence of white-lie telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48*(2), 160–181.

Lewis, M., Stanger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology, 25*(3), 439–443.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112*(3), 367–380.

Nurmsoo, E., & Robinson, E. J. (2009). Identifying unreliable informants: do children excuse past inaccuracy? *Developmental Science, 12*(1), 41–47.

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

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology, 43*(5), 1216–1226.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). New York: Cambridge University Press.

Peskin, J. (1992). Ruse and representations: on children's ability to conceal information. *Developmental Psychology, 28*(1), 84.

Rhodes, M., Gelman, S. A., & Brickman, D. (2010). Children's attention to sample composition in learning, teaching, and discovery. *Developmental Science, 13*(3), 421–429.

Shafto, P., Eaves, B., Navarro, D., & Perfors, A. (2012). Epistemic trust: modeling children's reasoning about others' knowledge and intent. *Developmental Science, 15*(3), 436–447.

Shafto, P., & Goodman, N. (2008). Teaching games: statistical sampling assumptions for learning in pedagogical situations. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society*.

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*(2), 669–679.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33*(3), 591–605.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search, volume 81 of lecture notes in statistics*. New York, NY: Springer.

Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development, 26*(5), 436–444.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. Kearns, S. A. Soller, T. K. Leen, & K. R. Muller (Eds.), *Advances in neural processing systems 11* (pp. 59–65). Cambridge: MIT Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*(7), 309–318.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Tanya, B., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–735.

Topál, J., Gergely, G., Miklósi, A., Erdohegyi, A., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science, 321*(September), 1831.

VanderBorght, M. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development, 71*(November 2008), 61–71.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655–684.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.

Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*(3), 288–297.