The Better Part of Not Knowing: Virtuous Ignorance

Jonathan F. Kominsky, Philip Langthorne, and Frank C. Keil

Yale University

Running Head: Virtuous Ignorance

Address for reprints and correspondence	:	Jonathan F. Kominsky Department of Psychology Yale University Box 208205 New Haven, CT 06520-8205
Email	:	jonathan.kominsky@yale.edu
Phone/Fax	:	(617) 877-4412/(203) 432-7172

Word Count	:	9784

Abstract

Suppose you are presented with two informants who have provided answers to the same question. One provides a precise and confident answer, and the other says that they do not know. If you were asked which of these two informants was more of an expert, intuitively you would select the informant who provided the certain answer over the ignorant informant. However, for cases in which precise information is practically or actually unknowable (e.g., the number of leaves on all the trees in the world), certainty and confidence indicate a lack of competence, while expressions of ignorance may indicate greater expertise. In three experiments, we investigated whether children and adults are able to use this "virtuous ignorance" as a cue to expertise. Experiment 1 found that adults and children older than 9 years selected confident informants for knowable information and ignorant informants for unknowable information. However, 5-6-year-olds overwhelmingly favored the confident informant, even when such certainty was completely implausible. In Experiment 2 we replicated the results of Experiment 1 with a new set of items focused on predictions about the future, rather than numerical information. In Experiment 3, we demonstrated that 5-8-year-olds and adults are both able to distinguish between knowable and unknowable items when asked how difficult the information would be to acquire, but those same children failed to reject the precise and confident informant for unknowable items. We suggest that children have difficulty integrating information about the knowability of particular facts into their evaluations of expertise.

Keywords: cognitive development; credibility; informants; confidence; epistemological beliefs

Sometimes the most impressive intellectual achievement can be recognizing the boundaries of one's own knowledge, or knowledge more generally. This is hardly a new idea, indeed it is one of the most classic philosophical themes across a wide range of cultures, whether it be Socrates ("…I am wiser than he is to this small extent, that I do not think I know what I do not know"[Plato, Apology 21d, tr. Tredennick, 1951]), Confucius ("Real knowledge is to know the extent of one's ignorance."[from Dunning, Johnson, Erlinger, & Kruger, 2003]), or Jefferson ("The wise know too well their weakness to assume infallibility; and he who knows most knows best how little he knows." [Jefferson, 1813]).

Knowing what one does not know may require considerable sophistication. To do so, more than just knowing what portion of available knowledge one possesses, one must have some sense of the full extent of what is potentially knowable, as well as what may never be knowable but still relevant. Indeed, in many (but not all) contexts, more knowledgeable individuals often have more accurate senses of their abilities and limitations, including of their knowledge and explanatory understandings, while less knowledgeable individuals tend to be miscalibrated and overconfident (Dunning, 2012; Dunning et al., 2003; cf. Arnold, Willoughby, & Calkins, 1985). Children are particularly severely miscalibrated about their own understanding of various phenomena (Mills & Keil, 2005).

Here, we turn to a different, but closely related, problem: Identifying when others are being overconfident about their own knowledge. In particular, we ask whether children and adults are capable of recognizing when an informant who says that they do *not* know the answer to a question is actually more knowledgeable than one who provides a confident and precise answer.

Virtuous ignorance

The literature on self-assessment suggests that claims of detailed knowledge are not invariably signs of expertise, and in some cases may be signs of incompetence or ignorance. In fact, under some circumstances claims of ignorance may indicate expertise and knowledge. For some questions or problems, the more expert individual may better understand how certain forms of information may be unknowable. For example, a novice might feel quite confident that one can know both the exact position and exact velocity vector of a particle at a given point in time, but an expert familiar with Heisenberg's Uncertainty Principle would freely admit that they could not know both precisely, and recognize any claims that one did as indicating ignorance rather than knowledge. Although sometimes satirized in terms of Donald Rusmfeld's talk of "known unknowns" (Rumsfeld, 2002), a sense of one's limits has traditionally been seen as a sign of insight and wisdom. We therefore distinguish two different expressions of ignorance: "Mere ignorance", the simple case of not knowing information that is, in principle, knowable, and "virtuous ignorance," which is an admission of not knowing something because the knowledge is impossible or implausible to obtain.

We expect that adults can identify many of these cases in which virtuous ignorance is a marker of greater expertise. The ability to know that one does not know (or *cannot* know) a specific piece of information often arises from a sophisticated understanding of the physical world and how uncertainty, chaotic systems, causal complexity, and cognitive demands make certain forms of precision or predictions highly implausible, if not impossible. Thus, one can evaluate the credibility of another's testimony in terms of how it meshes with one's own understanding of the plausibility of acquiring the attested knowledge. In such cases, the credibility of an informant may be especially in doubt when that individual expresses great certainty about information for which certainty is highly inappropriate. This facet of assessing confidence has been largely neglected in empirical studies with adults, but in extreme cases, it seems intuitively clear that adults will certainly reject confident declarations about logistically unknowable things. For example, it makes no sense for someone to claim that they know exactly, with great precision, how many leaves there are in all the trees in the world, and anyone who makes such claim with confidence should be regarded with great skepticism. However, it would be perfectly reasonable for someone to say that they did not know this information, and that this information could not plausibly be known.

Developmental challenges

Children, however, may have great difficulty rejecting a confident and precise answer, even when that answer is highly implausible. There are two key challenges children must confront. The first is an epistemological challenge: Children must recognize that the information is implausible or impossible to possess. To know that it is not feasible to have a precise number for all the leaves on all the trees in the world, one needs to have a sense of the immensity of the number, of the challenges of getting a snapshot of all leaves at a moment in time, and even of the ambiguities of when a budding or decaying leaf becomes or is no longer a leaf. Therefore, young children might fail to recognize virtuous ignorance as a cue to expertise because they fail to understand that possessing the information is implausible, and therefore favor almost any information they are given. While there are presumably some forms of knowledge that young children do recognize as impossible to possess, there are likely to be far fewer such cases for them than for adults.

The second challenge is that even in domains where children recognize that possessing some knowledge is implausible or impossible, they might find it nonetheless difficult to reject a confident informant. Strong confidence may often trump doubts about the plausibility of actually having a certain piece of knowledge. Such a "confidence heuristic", however, does not mean that children simply accept everything they are told. On the contrary, the extensive literature on testimony reveals that even preschoolers take evaluative stances towards claims made by others, and will take into account many source characteristics. These attributes can include: a source's past record of accuracy, a source's departure from consensus view, a source's current mental states and access to information, and a source's apparent dependency on other sources (for a review of this literature, see Robinson & Einav, 2014). In short, children appreciate that different sources should be trusted to different extents well before the start of formal schooling (Harris, 2012).

Even though very young children have some ability to detect whether informants are overconfident in their own knowledge, testimony evaluation skills and trust assessment ability continue to develop well past the preschool years. Preschoolers are more prone to trust what they hear even when lies and other deceptive behavior may have been quite salient in a speaker's past (Heyman, Sritanyaratana, & Vanderbilt, 2013; Mascaro & Sperber, 2009). Younger children, perhaps because of a need to acquire information as rapidly as possible, may simply take most things on faith except in the face of egregious informant incompetence. As children get older, they may become increasingly cautious with this form of blind deference as they potentially encounter an increased incidence of malicious informants. Younger children are also much more swayed by warmth and irrelevant competence than other, presumably more useful clues such as informant age and cultural or experiential background (Landrum, Mills, & Johnston, 2013;

Koenig & Stephens, 2014). Similarly, some of the subtleties of motivated reasoning escape younger children and can make it harder for them to be efficient skeptics (Mills & Keil, 2005, 2008).

These diverse developmental patterns raise the question of how children might learn to evaluate a source's confidence about knowledge claims. Confidence is certainly one clue to the quality of a source's knowledge, and one that children use. A sensitivity to confidence in demonstrations of object use emerges early in development, increasing dramatically during the second year of life (Brosseau-Liard & Poulin-Dubois, 2014). Even before rich language processing is available, non-verbal confidence is monitored as a way of judging the informativeness of others' actions. As language becomes the primary medium of communication, across a wide range of languages, young speakers are sensitive to linguistic indicators of certainty and confidence (Jaswal & Malone, 2007; Matsui, 2014), and by the time they enter the early school years, they are quite sophisticated evaluators of certainty-related expressions (Moore, Bryant, & Furrow, 1989) as well as being adept at indicating certainty in their own utterances (O'Neill & Atance, 2000).

Under ideal Gricean conditions, confidence presumably represents a source's genuine beliefs about what they know. Yet, even sincere sources might be mistaken and, of course, they could instead be blustering, bragging, or even lying about their own knowledge. Thus, a sophisticated evaluative stance towards confident declarations involves not only weighing the confidence message itself in terms of strength and specificity but also taking into account both source characteristics and the relationship between the confidence expressed and the content of the message. Taken together this is a complex integrative challenge. Young children might therefore over-value confidence early on as a marker of source information quality, perhaps at the expense of other factors. Confidence might be the simplest readily available heuristic for young children to use. Furthermore, confidence may be a more informative cue for younger children, as they might be genuinely less likely to encounter blustering, boastful, or deceptive adults.

In particular, we argue that 5-6-year-old children are heavily influenced by a person's confidence but have difficulty, relative to older children and adults, in calibrating informants (i.e., determining whether a person's confidence is diagnostic of their credibility), and rejecting poorly calibrated informants (Tenney, Small, Kondrad, Jaswal, & Spellman, 2011). Thus, when children and adults heard an account of an event from a confident expert and a cautious expert, and then saw that both experts were inaccurate, they reacted differently: While adults then favored the cautious expert on a subsequent task, 5-6-year-old children continued to favor the confident expert, despite understanding that the confident expert had been inaccurate. Similarly, we expect that children will continue to favor a confident expert even if they recognize that some information *could not* be accurate, because it is unknowable. However, there have been no explorations of whether children can understand that some information is unknowable in the first place and if they do, how they can use the idea of unknowability in subsequent judgments.

Predictions

Based on these observations, an overarching prediction emerges: Young children will favor an implausibly confident informant over a virtuously ignorant one. Put differently, expressions of virtuous ignorance can be a marker of being more credible, but one that young children may have great difficulty using in assessing an informant, particularly a confident informant. Supporting this developmental prediction, previous work has found that, through age

5, children are more likely to reject a (merely) ignorant informant than a blatantly inaccurate one (Mills et al., 2011).

Furthermore, we have discussed two challenges that could lead children to fail at this task: epistemological challenges and challenges integrating epistemology with confidence. These two accounts are not contradictory. In some domains children may fail to understand the implausibility of certainty; but we suggest that even when they do, the draw of confidence may be too strong to overcome.

We test these predictions with three experiments. In Experiments 1, we demonstrate that young children do indeed favor an implausibly confident informant despite glaringly unrealistic numerical precision, while older children and adults recognize when a virtuously ignorant informant is in fact a better expert, as they recognize the limits of what can be known. In Experiment 2, we demonstrate the same developmental shift for a completely different kind of implausibility involving predictions about future states of affairs, thereby showing the general dominance of the confidence heuristic in younger children. In Experiment 3, we ask children and adults to evaluate how difficult it would be to possess certain types of knowledge, and demonstrate that even when children can recognize the difficulty of knowing specific pieces of information, they tend to favor confidence over virtuous ignorance.

Experiment 1: Numerical Uncertainty

Experiment 1 examined how children come to appreciate implausibly precise claims of numerical knowledge. We expected that younger children, (i.e., 6 years and younger), would favor an expert who offered a precise answer to a question, even for items that adults would recognize as implausible. We base this prediction on prior work showing that children in this age group are both sensitive to certainty and confidence markers (Moore et al., 1989) and are also limited in their skepticism concerning expert sources (Mills & Keil, 2005). We studied children ages 5-10, expecting that by age 10 children would have achieved near adult mastery this task, as they are appropriately cynical about expert informants (Mills & Keil, 2005) and likely have a greater understanding of the implausibility of knowing certain things.

Methods

Participants. 105 children aged 5 to 10 years were divided into 4 grade clusters. Specifically, 26 children in Kindergarten (M_{age} = 67.1 mos., *SD* = 5.57 mos., 14 females), 27 children in first grade (M_{age} = 77.7 mos., *SD* = 4.38 mos., 14 females), 20 children in grades 2 and 3 (M_{age} = 98.5 mos., *SD* = 8.9 mos., 9 females), and 41 children in grades 4, 5, and 6 (M_{age} = 126.5 mos., *SD* = 8.9 mos., 17 females) participated in Experiment 1. Children were recruited in three ways: from regional schools, from regional science museums, and through visits into the experimenters' laboratory. All age groups were recruited by all three methods. Demographically, the children mirrored the demographics of the Connecticut population, which is approximately 82% white, 11 % African American, .5% Native American, 4 % Asian, and 2% two or more race and with 14% of Hispanic or Latino heritage. In addition, 53 adults were recruited on Amazon's Mechanical Turk System (M_{age} = 38.6 years, *SD* = 14 years, 41 females). An additional eleven adults and three first-graders participated in the experiment but were excluded for failing to complete the task.

Materials & Procedure. Adult participants viewed the stimuli through a web browser from their home computers and responded by clicking on a silhouette representing the expert they felt was “better” (see below). All child participants viewed the stimuli on Apple iPads and touched the screen to make their response. Children in grades K-2 had the text read aloud to them by the experimenter across all of the trials, whereas children in fourth grade were read aloud the instructions and the first item by the experimenter, who then allowed the children to read and advance through the remaining items themselves.

The experimenter explained to the child participants that they were about to play a detective game involving experts, and asked if they knew what the word “expert” meant. The experimenter then defined the word for children who did not know it, or redirected the definition provided by children who were able to generate one, to the following (based on similar training used by Keil, Stein, Webb, Billings, & Rozenblit, 2008):

> For this game, when we say “expert,” we mean “someone who really understands something really well.” So, someone could be an “expert” in basketball if they really understand basketball and how it works really, really well.

Children were then asked if they understood something really well and so could be an expert in that thing, to ensure comprehension. Next, the rules of the game were explained:

> We had a list of questions that we wanted to find the answers to, and so we found people who told us that they were “experts” about our topics and asked them our questions. You will see the questions that we asked them and the answers that they gave. For each pair of experts, please tap on the person that, in your opinion, is actually the BETTER expert about the topic.

We asked for the “better” expert because did not wish to imply that one of the individuals was not an expert and was therefore being dishonest. Rather, we wanted to establish that both individuals had some expertise, and have children evaluate which of them was a superior expert, based on the definition we provided. The instructions were read aloud to all children. Adults, participating online, read the experiment instructions on a page after answering age and sex demographics questions and before the first item, and did not complete the interactive comprehension check of identifying something in which they could be experts and receiving feedback.

Participants then completed 16 items, eight knowable and eight unknowable, in random order (see Appendix A for a complete list of items). The stimuli were presented to children one at a time, along with color depictions to serve as reminders of the claim (see example in Fig. 1). At the bottom of each page was a small button to advance to the next item.

For each item, one of the two silhouetted respondents (the “confident” expert) answered the question at the top of the page with a precise integer and the word, “exactly.” The other, “virtuously ignorant” expert responded, “I don’t know because it is not possible to answer that question precisely.” In all items, the numbers provided by the confident experts were “sharp”, that is, they were not rounded numbers like “10,000”, which can imply imprecision or hyperbole (Kao, Wu, Bergman, & Goodman, 2014). Participants were asked, “which one do you think is the better expert?”

Responses were scored as either 0 or 1 for each item based on accuracy, defined as follows: A score of 1 on a given item indicated a judgment that the person expressing certainty for a knowable item was the better expert or that the person saying “I don’t know” for an unknowable item was the better expert.

Invalid trials (missing data or stray clicks outside of the broad target regions covering the figures and their responses) were excluded from analyses. (This was infrequent and only occurred on 3.5% of all trials.) Final scores for each participant were computed as the proportion of valid responses of each item type that were accurate.

Finally, scores from any participant who had no responses recorded for either of the two categories were excluded from analysis. Also, because of unforeseen technical problems in excluding previous participants via the Amazon Mechanical Turk system, adults who submitted responses in more than one experiment in this set of experiments were excluded from all experiments and replaced by novel participants until all adult responses in all experiments came from unique Mechanical Turk worker identifiers.

Results

We calculated an accuracy score for all items, based on the “correct” choice of either the expert who gave a precise response for the knowable items or an “I don’t know” response for the unknowable items. These scores were computed as a proportion of the items responded to that the participant answered correctly (to allow for a few missing items). Fig. 2 shows the results of Experiment 1 by item type and age group. There was a main effect of item type, with much higher accuracy overall for knowable (M = .90, SD = .16) than unknowable items (M = .60, SD = .42), $F(1,163) = 108.67$, $p < .001$, $\eta_p^2 = .400$. This main effect was qualified by a significant interaction between age group and item type, $F(4,163) = 10.592$, $p < .001$, $\eta_p^2 = .206$.

Further analyses revealed a main effect of age group for unknowable items, $F(4, 163) = 25.063$, $p < .001$, $\eta_p^2 = .381$. Adults ($M = .90$, $SD = .27$) were significantly more likely to choose the "I don't know" (virtuously ignorant) expert than kindergarteners ($M = .28$, $SD = .30$), first graders ($M = .27$, $SD = .37$), and second and third graders ($M = .47$, $SD = .43$), all *p*s < .001. (All *p*-values reported for pairwise comparisons use Bonferroni correction.) Adults did not differ from children in grades four through six ($M = .74$, $SD = .36$), $p = .28$. Among the younger age groups, children in grades four through six were significantly more accurate than kindergarteners, $p < .001$, children in first grade, $p < .001$, and children in second and third grade, $p = .039$. There were no other significant differences for unknowable items. Furthermore, for unknowable items, kindergarteners and first graders selected the confident informant significantly more often than chance, *p*s <= .003, while children in fourth through sixth grades and adults selected the virtuously ignorant informant significantly more often than chance, *p*s < .001. Children in second and third grade were indistinguishable from chance responding, $p = .76$, further highlighting the developmental shift between first and fourth grade for unknowable items.

There was also an effect of age group for knowable items, $F(4, 163) = 8.84$, $p < .001$, $\eta_p^2 = .178$, but as the interaction suggests, it followed a somewhat different pattern. Adults ($M = .995$, $SD = .02$) were significantly more likely to choose the confident expert than kindergarteners ($M = .80$, $SD = .23$), first graders ($M = .89$, $SD = .20$), and fourth through sixth graders ($M = .88$, $SD = .15$), *p*s <= .029. However, they were not significantly different from second and third graders ($M = .92$, $SD = .12$), $p > .5$. There were no other significant age differences. These results indicate a slight developmental improvement in accuracy on knowable items, but a much smaller and less consistent improvement than found for unknowable items. Furthermore, all age groups were well above chance accuracy for knowable items, *p*s < .001.

Discussion

The ability to evaluate confident statements changes dramatically in the early school years, with children in kindergarten and first grade strongly favoring implausibly confident informants, while fourth graders and adults roundly reject such individuals and identify those who express virtuous ignorance as more knowledgeable. These findings fit well with earlier work that show that children are more likely to reject ignorant than inaccurate informants at age 5 (Mills et al., 2011), and suggest that children become aware of the value of a virtuously ignorant informant relatively late in development. While there was also some developmental improvement for picking the confident and precise informant on knowable items, that improvement was much smaller than the shift for unknowable items.

Before examining whether children were aware of the epistemic challenges involved in these items, we wanted to rule out a simple and uninteresting heuristic that might have generated the same pattern of results: The precise response always had a number in it, and the other response did not. It is possible that children simply favored any answer with a number, without even considering how easy or difficult it would be to know a given piece of information. In addition, it is also possible that can recognize epistemic implausibility for other forms of information even if they cannot do so for numerical precision. To rule out these explanations, Experiment 2 used predictions about the future rather than numerical precision to verify that children do indeed have a broad preference for implausibly confident informants.

Experiment 2: Future Uncertainty

Experiment 1 demonstrated that younger children were swayed much more by confident declarations of precise knowledge than by claims of virtuous ignorance even when precise knowledge was wildly implausible. One critical question concerns whether the insight that emerges is related to understanding numerical tallies. Given that number concepts can develop considerably during the school years (e.g., Siegler & Booth, 2004), younger children's endorsement and older children's rejection of extremely large and precise numerical quantities may reflect an emerging mathematical understanding and not a broader understanding of virtuous ignorance. In addition, because the implausible items generally had larger numbers than the plausible ones, that difference could have served as a relatively shallow clue to implausibility, rather than a general epistemological understanding.

To address these concerns, Experiment 2 looked at plausible and implausible knowledge without numerical precision by instead focusing on specific future predictions vs. unknowable ones. For example, one can confidently predict that a rainbow seen on October 1, 2224 will have a red stripe on top, but one cannot confidently predict that the most popular boy's name on October 1, 2224 will be George (or any other name). These intuitions seem to be based on a sense of stable predictable regularities as opposed to unpredictable outcomes, and importantly do not depend on numerical tabulations. Furthermore, previous work has established that, by five years of age, children are capable of sophisticated reasoning about future events (Atance & Meltzoff, 2005; Atance & O'Neill, 2005), and therefore we should expect that every age range we examined will have a sense of what is or is not plausible in making predictions. If the results of Experiment 1 are indeed due to a preference for confidence over plausibility, we should find the same developmental shift using these prediction items. However, if Experiment 1 merely revealed some kind of number-based heuristic, then we should see no developmental effects.

We created two sets of items: one set for which it was implausible or impossible to make highly specific predictions, and a second set where it was both possible and plausible to make highly specific predictions, as judged by an independent group of adults. We examined these items with slightly narrower age groups informed by the results Experiment 1, to verify the robustness of these developmental findings. In particular, we opted to focus on children in kindergarten, first grade, second grade, and fourth grade. In Experiment 1, kindergarteners and first graders favored a confident expert on unknowable items, second graders (as a subset of second and third graders) were at chance on unknowable items but significantly different from older children and adults, and fourth graders (as a subset of fourth through sixth graders) favored the virtuously ignorant informant and were not different from adults. Thus, these four child age groups offered an efficient option for replicating the overall developmental trajectory found in Experiment 1.

Methods

Participants. 26 children in Kindergarten (M_{age} = 66.7 mos., SD = 4.7 mos., 15 females), 20 children in first grade (M_{age} = 77.1 mos., SD = 3.7 mos., 12 females), 24 children in second grade (M_{age} = 90.9 mos., SD = 5.7 mos., 14 females), and 25 children in fourth grade (M_{age} = 113.8 mos., SD = 3.4 mos., 12 females) participated in Experiment 2. Children were recruited in the same manner as Experiment 1 from the same sources, but none had participated in Experiment 1. In addition, 24 adults who did not participate in Experiment 1 were recruited from Amazon Mechanical Turk (M_{age} = 30.0 years, SD = 16.2 years, 11 females).

Materials & procedure. The stimuli consisted of 16 predictions about topics to which children could easily relate, such as the weather outside or the color or shape of pieces of fruit. The time frames of the predictions from approximately 20 years in the future to over 700 years. All of the items contained questions that could be easily answered about the present (i.e., “How long is the president’s wife’s hair,” or “What letter comes after A in the alphabet”), but some could be answered about some specific point in the future and others could not. For example, it would be very safe to predict that, in 50 years, the letter B will come after the letter A in the alphabet, whereas it would be foolish to insist on what the length of the president’s wife’s (or husband’s) hair will definitely be at that time. In a pilot experiment we asked adults to rate whether it was plausible to make these predictions (and a number of alternative predictions not included in the experiment) with precision. We then used the 8 items with the highest agreement in either direction (i.e., 8 items that adults agreed were predictable for the “knowable” items and 8 items that adults agreed were not predictable for the “unknowable” items). These items can be found in Appendix B.

The appearance of the items was similar to Experiment 1, with the target question occupying the top third of the page, an illustration to aid retention in the middle third, and two side-by-side silhouettes representing the “experts” (with a green or blue background to differentiate them), and their responses below occupying the bottom third. As in Experiment 1, the presentation order of the 16 items was randomized, and the exact and “I don’t know” responses were randomly assigned to either the green or the blue figure on each trial. The procedure and scoring used for Experiment 2 were identical to those of Experiment 1 for all adult and child participants.

Results

Three second-graders were omitted for failing to provide any valid responses to at least one category of items. As in Experiment 1, accuracy was calculated separately for knowable and unknowable items and analyzed as a proportion of valid trials of each type for which participants were accurate. There was a strong interaction between grade and item type, $F(4, 106) = 11.27$, $p < .001$, $\eta_p^2 = .298$. Further analyses found that there were main effects of grade for both knowable items, $F(4,109) = 3.25$, $p = .015$, $\eta_p^2 = .107$, and unknowable items, $F(4,106) = 13.80$, $p < .001$, $\eta_p^2 = .342$. As Fig. 3 suggests, the developmental effects from Experiment 1 were largely replicated for unknowable items. Adults and fourth graders differed significantly from all younger age groups, Bonferroni-corrected *p*s < .05, but not from each other, $p = .34$. There were no significant differences between kindergarteners, first graders, and second graders. In addition, kindergarteners, first graders, and second graders selected the implausibly confident informant more often than chance, *p*s < .001, adults selected the virtuously ignorant informant more often than chance, $p < .001$, and fourth graders were at chance, $p = .676$.

As for knowable items, adults ($M = .722$, $SD = .195$) were significantly *less* accurate than first graders ($M = .907$, $SD = .142$), $p < .05$. No other age differences were significant for knowable items. This difference may be surprising initially, but it may reflect the strength of the confidence bias in younger children relative to adults: because adults are less seduced by confidence, they may be wary of it even when certainty may be appropriate, while children favor a confident informant no matter what. As in Experiment 1, all age groups were well above chance accuracy for knowable items, *p*s < .001.

Discussion

Experiment 2 largely replicated the findings of Experiment 1 with a new set of items focused on prediction rather than precise numerical estimation. The only notable difference was that the developmental shift for these future prediction items seems to come a little later than it did for the numerical estimation items, with second graders performing below chance, and fourth graders at chance. However, this diminished performance could also occur because these age groups represent the lower range of their respective age groups in Experiment 1 (i.e., no third graders or fifth to sixth graders). In any case, the same age differences were found, with a significant improvement on unknowable items around fourth grade. In this case we found no developmental improvement in accuracy for knowable items, in fact there was a slight developmental deterioration instead. This decline may provide further evidence for a diminishing bias for confident informants in adults: even when certainty may be appropriate, adults are wary of it. In the context of future predictions, adults may feel that some things that seem very predictable at face value can still change. However, they are still much more likely to select a virtuously ignorant informant in cases where certainty is completely implausible.

Having established that the developmental change in the use of a confidence heuristic is not unique to numerical estimation, the next task is to tease apart why children fail to recognize virtuous ignorance. One account is that they fail to recognize the epistemic challenges of knowing certain pieces of information. Another is that they may recognize when information is unknowable, but are unable to integrate this with expressions of confidence to recognize when someone is overstating their knowledge. To investigate these possibilities, in Experiment 3 we asked children to rate how hard it would be to know a given piece of information, and then had them complete the expert choice task using related items.

Experiment 3: Knowability vs. Confidence

Experiments 1 and 2 document a clear developmental shift from evaluating expertise in terms of having certain and confident discrete knowledge no matter what other factors are involved, to modulating such judgments as a function of the feasibility of having such knowledge. That feasibility in turn is assessed in light of how plausible it is to have such precision given real world patterns. The question remains as to whether younger children are largely unaware of such patterns of "knowability" or whether they do grasp knowability quite well but are unable to use it in evaluating informants. In other words, could children who show a strong bias to pick the confident expert in all conditions nonetheless be aware of the (im)plausibility of their answers? Children might recognize the implausibility of certain kinds of knowledge but nonetheless be unable to use this information to reject poorly calibrated informants. The results of Experiment 1 provide some support for this possibility. If the responses are analyzed in terms of how frequently the confident expert was chosen (rather than accuracy), there is systematic variation by item type. In fact, kindergarteners in Experiment 1 were still marginally less likely to select the confident expert for unknowable items than knowable, $p = .06$, and first graders were significantly less likely to choose the confident expert for unknowable items, $p = .006$, despite both groups picking the confident expert more often than chance on these items. This pattern suggests some awareness of the distinction between knowability and unknowability, but an inability to use this cue effectively.

To test this prediction, we constructed the strongest possible test, in which we first asked young children how knowable particular things were, and then immediately afterwards had them complete analogous items in the same format as Experiment 1. For the difficulty ratings, we changed the items to have narrower scope than the items in Experiment 1 (e.g, counting the number of blades of grass in Central Park vs. in New York State) so that the items in the expert evaluation were not pure repetitions, but still close enough that a judgment of implausibility for the narrower item would strongly entail a judgment of implausibility for the expert evaluation item. We validated these stimuli with adult piloting in MTurk, to verify that participants felt the rating items were comparable but of narrower scope. We predicted that young children would identify some types of knowledge as unknowable, or much harder to know, but still be drawn to pick a certain expert.

Methods

Participants. 31 children age 5-6 (M_{age} = 69.7 mos., *SD* = 14.0 mos., 17 females), 24 children age 7-8 (M_{age} = 95.8 mos., *SD* = 7.5 mos., 12 females) and 40 adults (M_{age} = 31.8 years, *SD* = 10.1 years, 16 females) participated in Experiment 3. Children were again recruited from the same populations as Experiments 1 and 2 but had not participated in those experiments. Adults were again recruited from Amazon Mechanical Turk and had not participated in the previous experiments.

Materials & procedure. Experiment 3 had two parts, difficulty rating and expert evaluation. In the first part, participants first completed four training items in which they learned to use the rating scale by rating the size of four animals (squirrel, cat, cow, and horse) on a scale that went from “SMALL” to “BIG.” The scale had no visible numerical values. We planned to exclude from analysis participants who failed to give higher average ratings to the cow and horse over the squirrel and cat (as this indicated an inability to use the scalar response method), but all participants succeeded at this task.

Children then saw a new scale which depicted a person climbing a set of very low steps on the left side with the word “EASY” below the image and another climbing up a very steep cliff face on the right side with the word “HARD” below the image. This scale was used with six new items, and for each item participants were asked to rate “How hard would it be to count…”. These six items were modified from Experiment 1, but instead of presenting the exact item from Experiment 1 (e.g., “The number of windows on the White House”), the item’s scope was narrowed to represent a subset of the original (e.g., “The number of windows on the President’s office in the White House”). An image was presented along with the question to help focus children’s attention, but these images were different from the ones used in Experiment 1 (and in the second part of this experiment). For these rating items, participants were told, “Touch the screen on the blue line, just like before, to say whether it would be easy, hard, or somewhere in between to count [item].” The six rating items were presented in random order. The full list of rating items for Experiment 3 can be found in Appendix C.

The second part of the experiment was identical to the procedure from Experiment 1 (including training), but only using the six items from Experiment 1 that corresponded to the difficulty rating items. The order of presentation was independently randomized and not tied to the presentation of the rating items. Children had all screens read aloud to them by the experimenter. Adult participants on MTurk once again saw the exact same stimuli but through a web browser on their personal computer, and responded via mouse clicks. Both the rating and the test phases of this experiment were completed for all subjects in a single session.

Results

We analyzed the difficulty ratings in terms of the absolute x-coordinate of the recorded mouse-click on the scale (within a constrained y-coordinate range), yielding a scale from 1 to 900 (pixels) with lower numbers representing "easier" responses. We conducted separate analyses of the difficulty ratings and accuracy in the expert evaluation task. For difficulty ratings, there was a main effect of item type, $F(1, 92) = 322.7$, $p < .001$, $\eta_p^2 = .778$, as well as a significant interaction between item type and age group, $F(2, 92) = 37.83$, $p < .001$, $\eta_p^2 = .451$. There was no significant main effect of age group, $F(2, 92) = 1.927$, $p = .151$. The average ratings by age group can be found in Table 1. As Table 1 reports, all age groups gave significantly higher ratings to unknowable than knowable items, indicating that even children in the youngest group were able to distinguish "knowable" from "unknowable" information. Further analyses showed main effects of age group for both knowable items, $F(2, 92) = 29.46$, $p < .001$, $\eta_p^2 = .390$, and unknowable items, $F(2, 92) = 15.44$, $p < .001$, $\eta_p^2 = .251$, but as Table 1 shows, these age effects went in different directions. For knowable items, adults gave lower ratings than both 5-6-year-olds and 7-8-year-olds, Bonferroni-corrected pairwise comparisons *p*s < .001, but the younger age groups did not differ from each other, $p > .9$. For unknowable items, 7-8-year-olds gave higher difficulty ratings than 5-6-year-olds, $p = .039$, and adults had higher difficulty ratings than both 5-6-year-olds and 7-8-year-olds, *p*s <= .045. There seems to be some developmental improvement in the ability to recognize epistemological challenges, which is not unexpected. However, in every age group, there is a clear distinction between knowable and unknowable information, thereby raising the question of whether this distinction carries over to the evaluation of putative experts.

As for the expert evaluation task, there was once again a main effect of item type, $F(1, 92) = 50.30$, $p < .001$, $\eta_p^2 = .353$, and age group, $F(2, 92) = 52.72$, $p < .001$, $\eta_p^2 = .534$, as well as a significant interaction, $F(2, 92) = 6.71$, $p < .001$, $\eta_p^2 = .127$. As in previous experiments there were main effects of age group for both knowable items, $F(2,92) = 5.33$, $p < .001$, $\eta_p^2 = .104$), and unknowable items, $F(2, 92) = 26.62$, $p < .001$, $\eta_p^2 = .367$. As can be seen in Fig. 4, for unknowable items, adults ($M = .825$, $SD = .292$) differed significantly from both 5-6-year-olds ($M = .237$, $SD = .346$) and 7-8-year-olds ($M = .431$, $SD = .423$), *p*s < .001, but the younger age groups did not differ from each other, $p = .127$. However, 5-6-year-olds chose the confident expert significantly more often than chance, $p < .001$, while 7-8-year-olds were equally likely to select the confident or virtuously ignorant expert, $p = .43$, and adults chose the virtuously ignorant expert significantly more often than chance, $p < .001$, closely replicating Experiment 1. Thus, younger children once again failed to reject a confident and precise informant, despite recognizing the greater difficulty of knowing information that adults classify as unknowable.

With fewer age groups and items than previous experiments, the same pattern emerged for knowable items, with adults ($M = .967$, $SD = .101$) differing significantly from both 5-6-year-olds ($M = .806$, $SD = .269$), $p = .013$, and 7-8-year-olds ($M = .819$, $SD = .311$), $p = .043$. However, once again all age groups were well above chance for knowable items, *p*s < .001, and the significant interaction indicates that the age effect was much smaller for knowable items.

We have argued that children recognize that these items are unknowable, but do not use unknowability to reject poorly calibrated experts. However, while children distinguished knowable and unknowable items in their difficulty ratings, difficulty ratings for unknowable items increased with age. It is possible that children thought these items were more difficult, but not truly unknowable, and therefore one could argue that they are able to reject poorly calibrated informants for items that they actually recognize as unknowable. Under this account, difficulty ratings should be a better predictor of performance on the expert evaluation task than age, since if there are individual items that children rated as highly as adults, then they should reject them just as adults do. However, under a calibration account, difficulty ratings should not be a good predictor, since even on items with high difficulty ratings children should not be able to use this information to reject a confident expert.

To distinguish between these possibilities, we conducted a backwards stepwise regression of accuracy on unknowable items in the expert evaluation task against age group, average difficulty rating of unknowable items, and an interaction term. This stepwise regression started with the model that included all three factors and then determined whether each factor was a significant predictor (meaning its coefficient β was significantly different from 0) using a *t*-test, and if not, removing it from the model. This analysis used the generous cutoff of the *p*-value for determining if the coefficient was different from 0, specifically, a factor was only removed from the model if the *t*-test that its coefficient was different from zero yielded $p >= .1$, and the process was repeated until no factors fit this criterion. This regression identified age group as the only significant predictor, first removing the interaction term, $\beta = -.245$, $p = .61$, and then the average difficulty rating, $\beta = -.053$, $p = .57$, leaving only the age group as a significant predictor of performance, $\beta = .601$, $p < .001$. This result demonstrates that the developmental increase in difficulty ratings for unknowable items cannot explain the difference in performance on the expert evaluation task, and another factor related to age makes the primary contribution.

Discussion

Experiment 3 demonstrated that children distinguished between knowable and unknowable items. However, despite being able to distinguish knowable and unknowable items in difficulty ratings, children were still more inclined to select an implausibly certain informant over a virtuously ignorant one, and this could not be accounted for by their ratings of the difficulty of possessing this knowledge. These results suggest that the primary developmental shift is not in assessing epistemic challenges (though that ability also improves with age), but rather a growing ability to integrate information about knowability and confidence when evaluating experts.

General Discussion

Children have difficulty using their epistemological understanding to recognize when a person who speaks with confidence might not know what he or she is talking about. Children have an early-developing sensitivity to expressions of confidence and certainty, and can use these expressions to evaluate statements. Here we have shown that younger children are so swayed by confidence and precision that they do not take into account those cases where virtuous ignorance is the stronger indicator of expertise.

Experiment 1 showed that young school children choose experts who confidently claimed to have implausibly precise numerical knowledge whereas older children and adults clearly rejected such claims and chose the expert who professed virtuous ignorance.

Experiment 2 showed that the younger children's difficulties with rejecting an inappropriately confident expert were not due to factors unique to numerical information. They showed the same difficulties taking into account implausible future predictions that did not involve numbers.

It would not be all that surprising if younger children simply had difficulty understanding the cognitive logistics of doing such things as counting all the leaves in a state or being able to predict far into the future. A great deal of knowledge about the world and of the epistemic challenges of tracking the world is needed to realize why some forms of claimed knowledge and expertise are highly implausible. Experiment 3 did show some developmental improvement in the ability to recognize the implausibility of knowing particular facts with precision. However, difficulty ratings did not predict the likelihood of rejecting the confident expert, but age did, indicating that above and beyond the developing ability to recognize unknowability, children cannot use unknowability to reject poorly calibrated experts. That said, the ability of even 5-6-year-old children to distinguish between knowable and unknowable information reveals a sophisticated epistemological stance.

An executive processing account

Experiment 3 showed that age is the primary predictor of participant's ability to reject an inappropriately confident expert, but age is of course only a proxy for some cognitive ability that develops between age 5 and adulthood. Recent studies have suggested that children's difficulties with using calibration to evaluate informants may be due to executive processing limitations. Two types of executive processing have been discussed extensively in this context, and both (or a combination of the two) are plausible explanations for our results. One explanation focuses on integrative capacity, the ability to hold two contrary things in mind and consider one in light of the other. For example, when adults are placed under significant cognitive load, they will fail to integrate an expert's past inaccuracy with their expressions of confidence, and will therefore not recognize that they are poorly calibrated (Tenney et al., 2011). While Tenney and colleagues' experiment showed a failure to integrate confidence with directly observed inaccuracy, our results could be explained the same way, replacing observed inaccuracy with epistemic implausibility. While Tenney et al. (2011) did not examine children's executive processing capacity directly, a typical measure of executive processing capacity is Backwards Digit Span

(BDS) performance, which has been used, for example, to examine the role of executive processing in children's performance on a false belief task (Davis & Pratt, 1995). Notably, BDS performance continues to improve middle childhood and even into early adolescence, or in other words, through the age range we examined (Dempster, 1981). Thus, executive processing capacity is a plausible (but untested) correlate of the developmental trajectory found in our expert evaluation task.

A second explanation focuses on inhibiting a default bias to believe what you are told (Gilbert, 1991; Gilbert, Krull, & Malone, 1990). One study with younger children (2.5-3.5 years old) found that they often accepted obviously false adult testimony about an event that the child themself witnessed, but the likelihood of rejecting the false testimony is positively correlated with performance on a spatial inhibitory control task (Jaswal et al., 2014). A recent study with older children also found that 6-7-year-olds' likelihood of choosing a cautious expert was correlated with parental ratings of inhibitory control, and inversely related to impulsiveness (Boseovski & Thurman, 2014). Indeed, more broadly, degrees of skepticism seem to be related to levels of inhibitory control (Jaswal & Pérez-Edgar, 2014). Furthermore, inhibitory control improves substantially between the ages of 5 and 10 (Williams, Ponesse, Schachar, Logan, & Tannock, 1999), particularly in the ability to inhibit a default response (the stop-signal task).

Either of these executive processing accounts, or a combination of both, fit our results well. Younger children show a significant preference for the confident expert, which could either reflect a total inability to integrate plausibility with confidence due to a capacity limitation, or much greater difficulty inhibiting a default response to trust a confident informant. Children ages 7 to 8 (second and third grades) are less consistent (at chance in Experiments 1 and 3, below chance in Experiment 2), which could indicate that they have the processing capacity to integrate these different types information some of the time (or that some children have the capacity but others do not), or that they have this capacity but are inconsistently able to inhibit their response to a confident informant. Overall this age group never performed significantly better than the younger age groups, perhaps because these challenges compounded each other. As noted above, the developmental trajectories of these two executive processing abilities are also very similar (see also Jacques & Marcovitch, 2010), and both align well with the developmental improvement in the expert evaluation task, so neither explanation can be easily ruled out. However, for these same reasons, both accounts are strong candidate mechanisms for explaining our results, and they are not at all mutually exclusive.

These questions could be explored more directly in future studies following the model of previous work that investigated the relationship between executive function and theory of mind. In particular, Carlson, Moses, & Breton (2002) provide a framework for a future study, as they used a battery of executive processing tasks that measure capacity and inhibitory control, as well as several other aspects of executive processing, in a preschool population. A follow-up to the current work could use a similar battery over a broader age range but would replace a theory of mind task with an expert selection task. Such a study would give a clearer account of whether and how executive processing influences the ability to use virtuous ignorance as a cue to informant quality. This study would provide a useful sequel to the primary findings discovered here, the dissociation between the ability to recognize unknowable information and the ability to reject an inappropriately confident expert.

In addition to this executive processing account, improvements in the ability to recognize the implausibility of possessing some information may also make a contribution. While we found developmental changes in the difficulty assessments of unknowable items in Experiment 3, these difficulty ratings did not predict endorsement of virtuously ignorant experts. Nonetheless, in order for improvements in executive processing to make a contribution, children must be able to recognize that the knowledge is implausible in the first place. Thus, the developing endorsement of virtuous ignorance more broadly most likely reflects improved ability to recognize implausible knowledge as well as improved ability to make use of that information.

Alternative accounts

We did not examine children's executive processing directly in these experiments, so we cannot completely rule out alternative accounts of our results. Experiment 3 demonstrated that there must be more at work than just the ability to recognize something as unknowable, so the crucial question is what relevant ability is developing over middle childhood. While executive processing is one possibility, there are others. Perhaps children simply prefer any information to none, even if that information seems totally implausible. However, as mentioned in the introduction to Experiment 3, even though children favor the confident informant much more than adults and more often than chance, they still favor the confident informant slightly less on unknowable items compared to knowable items. A general preference for any information over none would suggest a more uniform bias across item types, while this pattern fits better with the idea that children are failing in their attempts to make use of epistemic plausibility in evaluating experts.

This observation does not contradict one further alternative account, that younger children's concept of an "expert" includes someone who possesses knowledge that would otherwise seem implausible to possess. There is some reason to expect that children are willing to accept unexpected or otherwise dubious information from an informant who shows clear communicative intent, under the assumption that they know something that the child does not (e.g., Jaswal, 2004). However, this account proposes that these effects are driven by qualities specific to the concept of an "expert", which precludes it from explaining previous findings in the trust and testimony literature. For example, children will favor a nice informant who did not have visual access to the relevant information over another informant who did have access to that information (Lane, Wellman, & Gelman, 2013). An account of our results based on a unique property of the concept of "expert" would require a different mechanism to account for Lane et al. (2013)'s results, since in their experiment neither informant was ever described as an "expert". In contrast, the executive processing account offers a more parsimonious explanation for both the current work and previous findings: The difficulties children have in this experiment reflect a broader challenge of using epistemological information to disbelieve an informant with positive (or even neutral) traits.

Finally, there is the possibility that children judging that this knowledge was difficult to possess led them to think better of the informants who possessed it, essentially assigning competence to them based on an epistemological achievement and therefore selecting them as the better expert. There are two issues with such an account. The first issue is that it also predicts a relationship between difficulty ratings and expert selection, but in the opposite direction of the one observed, at least for younger children. If this were the case, the stepwise regression in Experiment 3 would have produced an age x difficulty rating interaction, as younger children would presumably be more likely to select the confident expert for more difficult items, while in older children and adults the relationship would be inverted or nonexistent. The second issue is that, while younger children do seem to assign positive traits to individuals who acquire intelligence through effort, older children do so as well, and to the same degree (Lockhart, Keil,

& Aw, 2013; cf. Heyman, Gee, & Giles, 2003). Thus, this account cannot explain the developmental changes between younger and older children.

Identifying unknowability

Despite these failures to reject inappropriate confidence in favor of virtuous ignorance, it is encouraging to note that even five-year-olds in Experiment 3 distinguished between knowable and unknowable information. This suggests that, with some assistance using the information they already possess, these children might easily learn the value of virtuously ignorant informants. However, questions remain concerning how to determine whether something is or is not knowable. The process of plausibility determination is not well understood. Adults could easily be seduced by an overconfident informant if they did not understand that such confidence was implausible. For example, recalling the example from the introduction, an adult who did not know that it was impossible to simultaneously know both the position and velocity vector of a particle might favor a confident and precise "expert". In the reported experiments, we validated our stimuli with adults to ensure that the selected items were recognized as implausible to be confident and precise about, but in day-to-day life, it is unclear how well adults can actually identify the plausibility of knowing something.

This problem is compounded if we consider how such plausibility information might be learned. Given the intricate web of deference needed to successfully navigate a complex world (Keil et al., 2008), a lay sense of knowabilty may often come from the very experts that we are trying to evaluate. For example, of the readers of this paper who knew of Heisenberg's Uncertainty Principle, it is unlikely that any of them have direct evidence for it or proved it themselves. Indeed, the authors themselves only know it through deference to physics experts. If there were an equal population of experts that claimed that such information was knowable with precision, how would we be able to evaluate whether a confident, precise response was appropriate? To give a more everyday example, when a medical doctor or IT professional confidently claims to have identified the exact source of (and/or solution to) a problem, many laypeople (and many experts in other domains) would have difficulty evaluating the appropriateness of such confidence, while more knowledgeable individuals might feel that some problems are too complex to diagnose so straightforwardly.

It is also worth noting that our "virtuously ignorant" informants did not simply say "I don't know". They provided a specific reason for their ignorance, i.e., that the information could not be known. This is again a key distinction between "virtuous ignorance" and "mere ignorance". It is not necessarily a cue of expertise to merely express ignorance, even when something is in fact unknowable. One could claim ignorance because one is not an expert as easily as one could claim ignorance because one *is* an expert. It would be somewhat surprising if someone who expressed ignorance without providing further information would ever be seen as an expert. While not specifically tested in these experiments, we would expect that the additional statement that specific information is unknowable is important for identifying a virtuously ignorant expert over someone who simply knows nothing.

Conclusion

Young children seem to recognize when certain information is difficult or implausible to possess, but have great difficulty overcoming their bias to believe a confident informant over a virtuously ignorant one. To successfully identify true experts in the many areas where human knowledge is highly incomplete, children must develop the ability to reject an inappropriately confident informant based on their epistemic understanding of what knowledge can be feasibly possessed. However, for us to have a complete understanding of virtuous ignorance, further work is required to identify how and when children and adults conclude that something is unknowable. There may be cases where even adults, who can reject a confident expert when they recognize that such confidence is implausible, may not realize that they should do so because of a failure to recognize more subtle cases of epistemic feasibility.

References

Arnold, L., Willoughby, T. L., & Calkins, E. V. (1985). Self-evaluation in undergraduate medical education: A longitudinal perspective. *Journal of Medical Education*, *60*(1), 21-8.

Atance, C. M., & Meltzoff, A. N. (2005). My future self: Young children's ability to anticipate and explain future states. *Cognitive Development*, *20*, 341-361.

Atance, C. M., & O'Neill, D. K. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation*, *36*(2), 126-144. doi:10.1016/j.lmot.2005.02.003

Boseovski, J. J., & Thurman, S. L. (2014). Evaluating and approaching a strange animal: Children's trust in informant testimony. *Child Development*, *85*(2), 824-34. doi:10.1111/cdev.12156

Brosseau-Liard, P. E., & Poulin-Dubois, D. (2014). Sensitivity to Confidence Cues Increases during the Second Year of Life. *Infancy*, *19*(5), 461-475.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73-92. doi:10.1002/icd.298

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology*, *47*(1), 25-31. doi:10.1080/00049539508258765

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, *89*(1), 63. doi:10.1037/0033-2909.89.1.63

Dunning, D. (2012). *Self-insight: Roadblocks and detours on the path to knowing thyself*. Psychology Press.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83-87.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107. doi:10.1037/0003-066X.46.2.107

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*(4), 601. doi:10.1037/0022-3514.59.4.601

Harris, P. L. (2012). Trusting what you're told: How children learn from others. Cambridge; Harvard University Press.

Heyman, G. D., Sritanyaratana, L., & Vanderbilt, K. E. (2013). Young children's trust in overtly misleading advice. *Cognitive Science*, *37*(4), 646-667. doi:10.1111/cogs.12020

Heyman, G. D., Gee, C. L., & Giles, J. W. (2003). Preschool children's reasoning about ability. *Child Development*, 516-534.

Jacques, S., & Marcovitch, S. (2010). Development of executive function across the life span. In W. F. Overton (Ed.), *The handbook of life-span development: Vol I: Cognition, biology, and methods across the lifespan* (pp. 431-466). Hoboken, NJ: Wiley.

Jaswal, V. K. (2004). Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development*, *75*(6), 1871-85. doi:10.1111/j.1467-8624.2004.00822.x

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-Year-Olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development*, *8*(3), 263–283. doi:10.1080/15248370701446392

Jaswal, V. K. & Pérez-Edgar, K. (2014). Resolving conflicts between observation and testimony: The role of inhibitory control. In E. J. Robinson & S. Einav (Eds.), Current Issues in Developmental Psychology: Children's Trust in Testimony. New York: Psychology Press.

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, Online First. doi:10.1111/desc.12187

Jefferson,T. (1813) in a letter to Dr. Benjamin Rush, March 6, 1813, in Gordon Carruth and Eugene Erlich, the Harper Book of American Quotations (New York: Harper & Row, 1988), p. 256.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(33), 12002-7. doi:10.1073/pnas.1407479111

Keil, F. C., Stein, C., Webb, L., Billings, V. D., & Rozenblit, L. (2008). Discerning the division of cognitive labor: An emerging understanding of how knowledge is clustered in other minds. *Cognitive Science*, *32*(2), 259-300. doi:10.1080/03640210701863339

Koenig, M. A., & Stephens, E., (2014). Characterizing Children's Responsiveness to Cues of Speaker Trustworthiness: Two Proposals. In E. Robinson and S. Einav (Eds.), *Trust and Skepticism: Children's Selective Learning from Testimony*. Psychology Press: Cambridge, UK

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*(4), 622-38.

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*(4), 1253-68. doi:10.1111/cdev.12029

Legare, C. H., Mills, C. M., Souza, A. L., Plummer, L. E., & Yasskin, R. (2013). The use of questions as problem-solving strategies during early childhood. *Journal of Experimental Child Psychology, 114*(1), 63-76.

Lockhart, K. L., Keil, F. C., & Aw, J. (2013). A bias for the natural? Children's beliefs about traits acquired through effort, bribes, or medicine. *Developmental Psychology*, *49*(9), 1669-82. doi:10.1037/a0030769

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

Matsui, T. (2014). Children's understanding of linguistic expressions of certainty and evidentiality. *Pragmatic Development in First Language Acquisition*, *10*, 295.

Mills, C. M., Legare, C. H., Grant. M. G., & Landrum, A. R. (2011). Determining whom to question, what to ask, and how much information to ask for: The development of inquiry in young children. *Journal of Experimental Child Psychology, 110,* 539-560.

Mills, C. M., & Keil, F. C. (2005). The development of cynicism. Psychological Science, 16, 385-390

Mills, C. M., & Keil, F. C. (2008). The development of (im)partiality. Cognition, 107, 528-551.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, *60*, 167–171.

O'Neill, D. K., & Atance, C. M. (2000). "Maybe my daddy give me a big piano": The development of children's use of modals to express uncertainty. *First Language*, 20, 29–52.

Plato, & Tredennick, H. (1954). *The Last Days of Socrates. The Apology, Crito, Phaedo. Translated and with an Introduction by Hugh Tredennick*. Penguin Books.

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

Rumsfeld, D. (2002). *DoD News Briefing - Secretary Rumsfeld and Gen. Myers, February 12, 2002 11:30 AM EDT.* U.S. Dept. of Defense, Washington, D.C., retrieved from http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*(2), 428-444.

Tenney, E. R., Small, J. E., Kondrad, R. L., Jaswal, V. K., & Spellman, B. A. (2011). Accuracy, confidence, and calibration: How young children and adults assess credibility. *Developmental Psychology, 47*, 1065-1077.

Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology*, *35*(1), 205-213. doi:10.1037/0012-1649.35.1.205

Acknowledgements

This research was supported by National Institutes of Health Grant R-37-HD023922 to Frank C. Keil. The authors would like to thank Kasia Hitczenko, Anh Ha, and Julia Merriam for their assistance with stimulus design and data collection.

Table 1: Difficulty ratings for unknowable and knowable items in Experiment 3.

Age Group	Avg. Unknowable	Avg. Knowable	Paired-samples *t*-test
5-6-year-olds	556.5 (163.7)	371.4 (129.9)	$t(30) = 4.15$, $p <.001$
7-8-year-olds	648.2 (137.7)	342.6 (125.2)	$t(23) = 10.744$, $p < .001$
Adults	733.5 (100.6)	172.2 (102.5)	$t(39) = 23.464$, $p < .001$

NOTE: numbers in parentheses are standard deviations.

Figure 1. Example stimuli from Experiment 1.

Figure 2. Results of Experiment 1. While there were effects of age for both item types, the effects on unknowable items were much more dramatic and showed a clear developmental shift between first and fourth grades. Error bars represent 95% CIs.

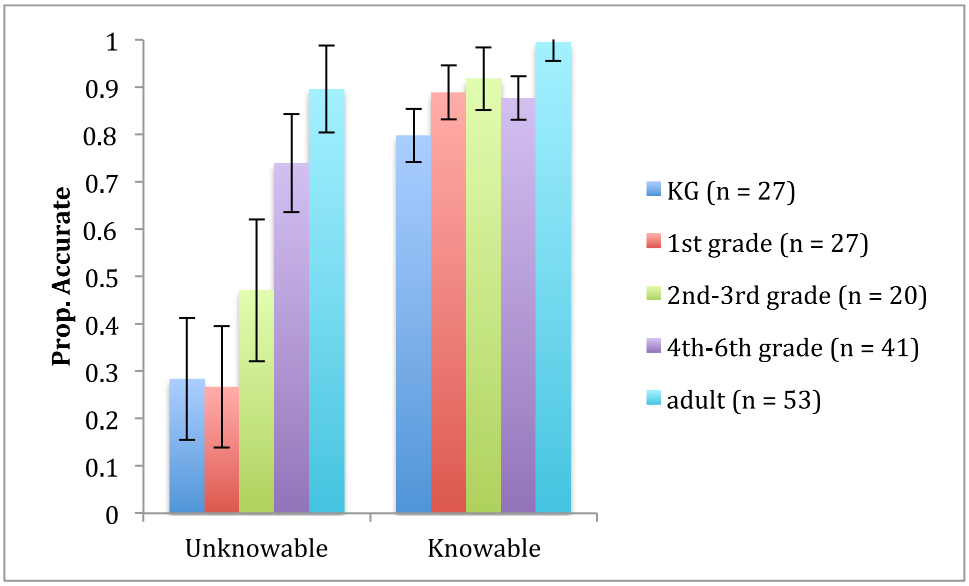


Figure 3. Results of Experiment 2. Error bars represent 95% CIs.

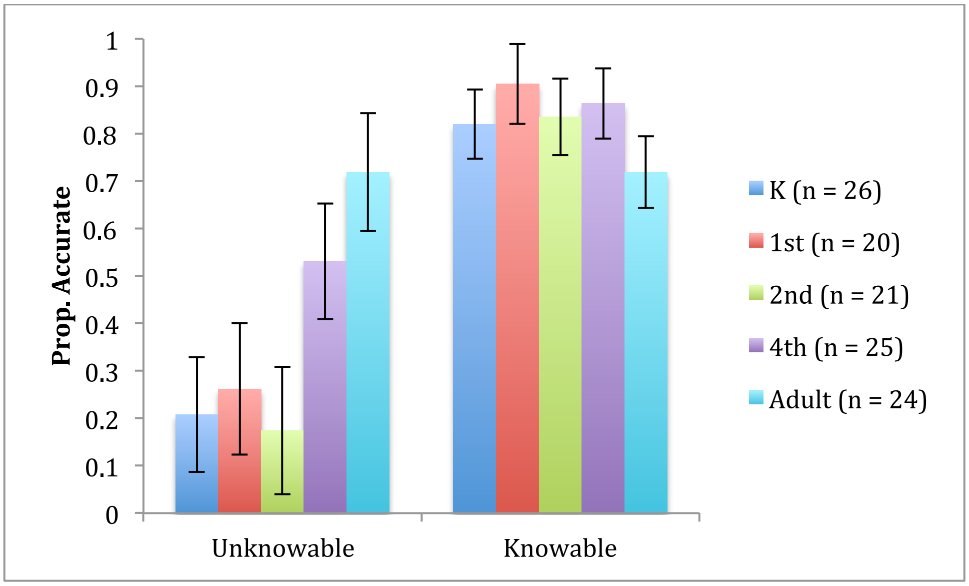
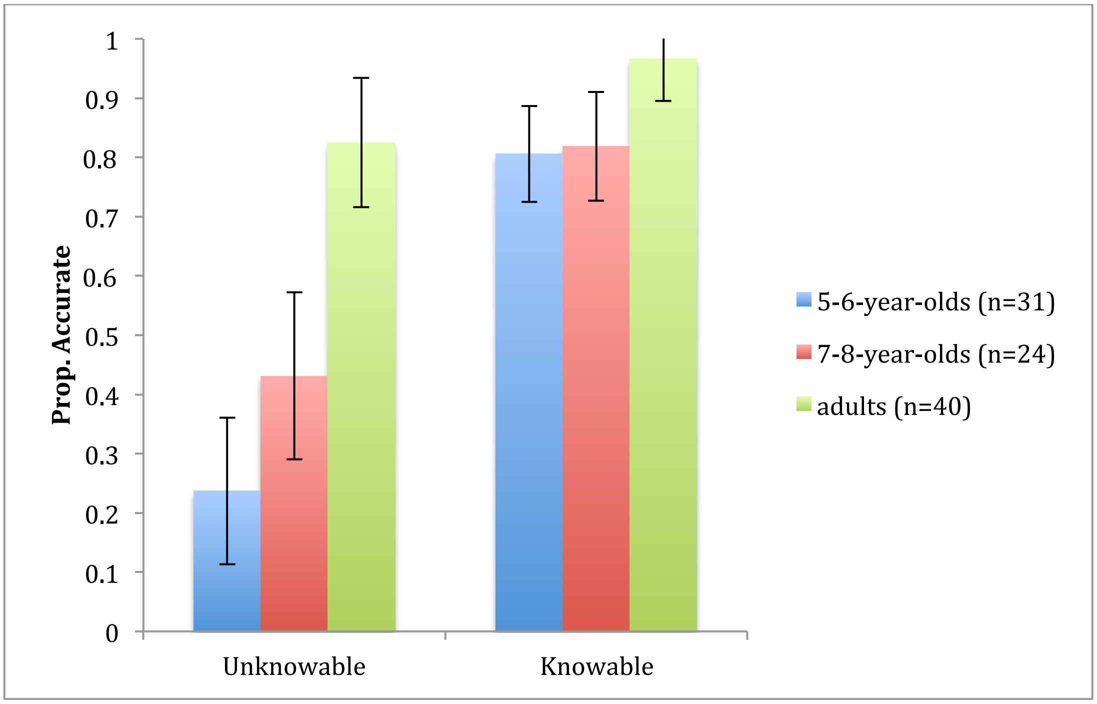


Figure 4. Results of the expert evaluation task in Experiment 3. Error bars represent 95% CIs.



Appendix A: Stimuli used in Experiment 1

IDK response for all: I don't know because it is not possible to answer that question precisely.

1. If you count the number of windows on the White House, how many will you get?
   ⇒ There are exactly 147 windows on the White House.
2. If you count the number of major islands in Hawaii, how many will you get?
   ⇒ There are exactly 8 major islands in Hawaii.
3. If you count the number of keys that are normally on a piano, how many will you get?
   ⇒ There are exactly 88 keys on a piano.
4. If you count the number of fins that are normally on a tiger shark, how many will you get?
   ⇒ There are exactly 8 fins on a tiger shark.
5. If you count the number of strings that are normally on a harp, how many will you get?
   ⇒ There are exactly 47 strings on a harp.
6. If you measure the length of a dollar bill in inches, how many will you get?
   ⇒ There are exactly 2.61 inches across a dollar bill.
7. If you count the number of wings that are normally on a dragonfly, how many will you get?
   ⇒ There are exactly 4 wings on a dragonfly.
8. If you count all the bones that are normally in a rabbit, how many will you get?
   ⇒ There are exactly 206 bones in a rabbit's body.
9. If you count all the leaves on all trees in the entire world, how many will you get?
   ⇒ There are exactly 809,343,573,353,235 leaves on all trees in the world.
10. If you count the number of sticks of chalk that have ever been used in all schools in the world in the past ten years, how many will you get?
    ⇒ In the past ten years, exactly 224,463,723 sticks of chalk were used in all schools in the world.
11. If you count the number of times all ballerinas in the world jumped last year, how many will you get?
    ⇒ Last year, all ballerinas in the world jumped exactly 30,975,224 times.
12. If you count the number of cars with cracked windshields everywhere in the world last year, how many will you get?
    ⇒ There were exactly 98,351,575 cracked car windwhields in the world last year.
13. If you count the number of seagulls that landed on all beaches in all the world in the past year, how many will you get?
    ⇒ Exactly 17,452,754 seagull landed on all beaches in the world in the past year.
14. If you count the number of blades of grass that sprouted in New York state last year, how many will you get?
    ⇒ In the last year, exactly 537,454,265,729,986,534 blades of grass sprouted in New York state.
15. If you count all the flies that were eaten by spiders in the last year in all of Connecticut, how many will you get?
    ⇒ In the last year, exactly 39,343 flies were eaten by spiders in all of Connecticut.
16. If you count all the people who rode in elevators in the last year in the whole world, how many will you get?
    ⇒ In the last year, exactly 4,934,524,643 people rode in elevators in the whole world.

Appendix B: Stimuli used in Experiment 2

IDK response for all: I don't know because it is not possible to answer that question precisely.

1. What season will it be in the United States on January 24, 2064?
   ⇒ It will definitely be winter in the United States on January 24, 2064.
2. What shape will oranges be on September 14, 2032?
   ⇒ Oranges will definitely be round on September 14, 2032.
3. Will pencil marks be erasable on December 16, 2032?
   ⇒ Pencil marks will definitely be erasable on December 16, 2032.
4. What colors will a rainbow have on April 4, 2721?
   ⇒ A rainbow will definitely have the colors red, orange, yellow, green, blue, indigo, and violet on April 4, 2721.
5. Will sugar taste sweet on November 4, 2098?
   ⇒ Sugar will definitely taste sweet on November 4, 2098.
6. What color stripes will zebras have on August 4, 2090?
   ⇒ Zebras will definitely have black stripes on August 4, 2090.
7. What color will lemons be on September 14, 2032?
   ⇒ Lemons will definitely be yellow on September 14, 2032.
8. What letter will come after A in the alphabet on October 29, 2084?
   ⇒ The letter B will definitely come after A in the alphabet on October 29, 2084.
9. How long will the president's spouse's hair be, in inches, on February 17, 2033?
   ⇒ The president's spouse's hair will definitely be 15 inches long on February 17, 2033.
10. Which city will have the most bike accidents on October 10, 2312?
    ⇒ Cincinnati will definitely have the most bike accidents on October 10, 2312.
11. What will be the high temperature on November 24, 2144 in Buffalo, NY?
    ⇒ The high temperature will definitely be 52 Fahrenheit on November 24, 2144 in Buffalo, NY.
12. What will be the most popular boys' name on October 22, 2322?
    ⇒ The most popular boys' name on October 22, 2322 will definitely be Blaise.
13. When will the first earthquake be in San Francisco after September 14, 2213?
    ⇒ The first earthquake in San Francisco after September 14, 2213 will definitely be on December 29, 2213.
14. What movie will make the most money on December 21, 2100?
    ⇒ The movie "Journey to Expedia" will definitely make the most money on December 21, 2100.
15. Who will have the most popular song on the radio on March 4, 2234?
    ⇒ Annabelle Friedman will definitely have the most popular song on the radio on March 4, 2234.
16. What will be the name of the next planet in the whole universe that will be discovered after October 7, 2533?
    ⇒ The name of the next planet that will be discovered in the whole universe after October 7, 2533 will definitely be Echnidna.

Appendix C: Stimuli used in Experiment 3

Training Questions:
Is this animal the size of a mouse, the size of an elephant, or somewhere in between?
1. Squirrel
2. Cat
3. Cow
4. Horse

Initial Difficulty Rating Questions:
How easy would it be to count the number of grey seagulls that landed on all the beaches in Florida last year?
How easy would it be to count the number of blades of bluegrass that sprouted in Central Park last year?
How easy would it be to count the number of sticks of chalk that were used in all the schools in Los Angeles, California in the last 10 years?
How easy would it be to count the number of windows in the President's office in the White House?
How easy would it be to count the number of black keys on a piano?
How easy would it be to count the number of wings on green dragonflies?

Target Questions (from Experiment 1, see Appendix A):
If you count the number of windows on the White House, how many will you get?
If you count the number of keys that are normally on a piano, how many will you get?
If you count the number of wings that are normally on a dragonfly, how many will you get?
If you count the number of sticks of chalk that have ever been used in all schools in all the world in the past ten years, how many will you get?
If you count the number of seagulls that landed on all beaches in all the world in the past year, how many will you get?
If you take the number of blades of grass that sprouted in New York state last year, how many will you get?