Journal of Experimental Child Psychology 176 (2018) 73–83

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Children's evaluation of verified and unverified claims

Lucas Payne Butler [a,*], Marco F.H. Schmidt [b,c], Nadia S. Tavassolie [a], Hailey M. Gibbs [a]

[a] Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742, USA
[b] International Research Group "Developmental Origins of Human Normativity", Ludwig-Maximillians-Universität, Munich 80539, Germany
[c] Department of Developmental and Educational Psychology, University of Bremen, Bremen 28359, Germany

ARTICLE INFO

Article history:
Received 22 May 2018
Revised 2 July 2018
Available online 20 August 2018

Keywords:
Social cognition
Social learning
Epistemic cognition
Epistemic vigilance
Selective trust
Normative evaluation

ABSTRACT

Critical to children's learning is the ability to judiciously select what information to accept—to use as the basis for learning and inference—and what information to reject. This becomes especially difficult in a world increasingly inundated with information, where children must carefully reason about the process by which claims are made in order to acquire accurate knowledge. In two experiments, we investigated whether 3- to 7-year-old children (N = 120) understand that factual claims based on verified evidence are more acceptable than claims that have not been sufficiently verified. We found that even at preschool age, children evaluated verified claims as more acceptable than insufficiently verified claims, and that the extent to which they did so was related to their explicit understanding, as evident in their explanations of why those claims were more or less acceptable. These experiments lay the groundwork for an important line of research studying the roots and development of this foundational critical thinking skill.
© 2018 Elsevier Inc. All rights reserved.

Introduction

Much of children's learning about the world relies on what others tell them (Gelman, 2009; Harris & Koenig, 2006; Harris, 2012; Harris, Koenig, Corriveau, & Jaswal, 2018; Koenig, Clément, & Harris,

* Corresponding author.
E-mail address: lpbutler@umd.edu (L.P. Butler).

https://doi.org/10.1016/j.jecp.2018.07.007
0022-0965/© 2018 Elsevier Inc. All rights reserved.

2004). However, learning from testimony presents an epistemological challenge in that it relies almost entirely on our trust in the individual we hope to learn from—whether parents, teachers, mentors, or peers. To learn effectively from the testimony of others, children must first learn to balance trust with vigilance and to reject testimony that is false or based on insufficient evidence (Sperber et al., 2010).

Contemporary research in developmental psychology has shown that very young children are indeed epistemically vigilant. They use both social and epistemic cues—such as whether informants share an accent or group membership with them (Corriveau, Kinzler, & Harris, 2013), whether informants have a past history of accuracy (Clément, Koenig, & Harris, 2004), and whether informants seem to have access to evidence supporting their claims (Brosseau-Liard & Birch, 2011; Nurmsoo & Robinson, 2009)—in order to selectively choose which informants are most likely to be reliable in providing information.

Furthermore, there is a growing body of work supporting the idea that children have a developing capacity to identify and judge an informant's ignorance, incompetence, and/or inaccuracy (Mills, 2013) and may use this mental state reasoning to characterize individuals' relative trustworthiness (Koenig & Harris, 2005). Young children also discriminate between cases in which informants show a pattern of accuracy and cases in which they can truly be deemed knowledgeable (Einav & Robinson, 2011). Indeed, children are fairly adept at using both person-specific cues (e.g., history of prior accuracy) and situation-specific cues (e.g., current visual access) to judge who is likely to provide reliable information. They show some ability to flexibly use these different types of cues when judging different types of testimony based on the specific type of claim being made (Brosseau-Liard & Birch, 2011; Koenig et al., 2015). Furthermore, children recognize who has access to relevant information (Pillow, 1989) and use that to guide who to use as a source for information (Robinson, Butterfill, & Nurmsoo, 2011). Finally, they will excuse informants for inaccuracy in cases where they are unable to corroborate informants' claims (Koenig, 2012; Nurmsoo & Robinson, 2009).

Thus, children appear to be quite sophisticated in their selective judgments of who to trust. However, making judgments about who is a trustworthy informant is a necessary but not fully sufficient component of epistemic reasoning. Trait-like factors such as informants' prior accuracy or reliability should give us a high prior expectation that their claims are likely to be true. But this assumption of trustworthiness must have its limits. In many cases, even experts make claims about things they may know little about. Therefore, it remains important to evaluate the process by which they make those claims—even given a prior history of accuracy or inaccuracy. Children ought to be able to overlook prior trustworthiness as they reason about specific claims that are clearly verifiable.

As an illustrative example, a doctor may be highly educated and can generally be relied on to provide accurate information about biological processes and medical issues. However, in a situation where some evidence is not available to the doctor, say the contents of an unlabeled medicine bottle, the doctor's expertise and prior accuracy no longer have any relevance to the reliability of a claim she or he may make about the bottle's contents. The doctor must first *verify* the contents of the bottle before she or he can make a valid claim about them. If the doctor claims knowledge without sufficient verification, we should reject that claim even though it is being made by an expert. Even within the doctor's area of expertise, such education and training are not sufficient to support claims she or he cannot make without additional evidence. The doctor cannot diagnose cancer without a biopsy, HIV without a blood test, or hearing problems without a hearing screen. Although the accumulated knowledge of medical training is important, likely the most critical skill a medical expert learns is what evidence the expert needs to make a diagnosis and how that evidence ought to be gathered.

The current research addressed this specific question of how young children develop the ability to evaluate and reason about others' claims. To gain initial purchase on this issue, the current research investigated at what age children comprehend that factual claims need to be verified in advance and how this affects their evaluation and acceptance of novel claims. Specifically, we asked whether and when young children can evaluate claims on the basis of whether the individual making a claim has first verified it by generating sufficient and necessary evidence. We also traced the developmental trajectory of this important skill by investigating children at an age when, although they have likely had some experiences in formal education, they are unlikely to have had any formal training in this kind of reasoning. Children were introduced to a digital storybook in which characters encounter unusual containers and make factual claims about their contents. Crucially, the characters in the

storybooks made those claims on the basis of *full, insufficient,* or *no* verification. We then asked children a series of questions about the acceptability of those claims and prompted them to justify their evaluations.

Critically, this research was not about selective trust in informants but rather about the claims themselves. This research made two important contributions. First, it focused not on the accuracy or reliability of informants but rather directly on children's understanding of the evidentiary process underlying factual claims. Factual claims are not epistemic claims to knowledge (e.g., "I know 'X'") but rather assertive claims that a proposition is empirically true (e.g., simply asserting "X") regardless of whether they are accurate. This focus allowed us to isolate a key empirical reasoning capacity that, although implicit in much of the research reviewed above, has rarely, if ever, been investigated directly. Second, this work delves into the reasoning behind children's evaluations of others' claims and how developmental changes in that reasoning may relate to children's ability to differentially evaluate verified and unverified claims.

Experiment 1

In Experiment 1, we tested whether 4- to 7-year-old children would successfully judge unverified claims as less acceptable than verified ones. We had several key hypotheses. First, we predicted that children would be less accepting of unverified claims than of verified ones. Second, we predicted that if children truly understand the link between verification and factual claims, they should distinguish between sufficient and insufficient verification in their evaluations rather than simply between any verification and no verification at all. Third, we predicted that the extent to which children made this distinction would be related to their ability to explicitly justify their evaluations and that this would show developmental change over this age range.

Method

Participants

In Experiment 1, we tested 24 preschool children (mean age = 57.0 months, range = 48–70, 12 boys) and 24 early elementary school children (mean age = 86.5 months, range = 72–95, 12 boys). Children were recruited either at a local children's museum in a medium-sized, mid-Atlantic U.S. city, or into the lab at a large university in a nearby town. Although the sample was largely upper-middle class, children represented a range of ethnic backgrounds, with the sample being 45% White/Caucasian, 8.5% Black/African American, and 46.5% from other or mixed ethnic backgrounds.

Procedure

Children were read a storybook (presented as PowerPoint slides on a laptop). Children were first introduced to three puppets and told that they were going to see the puppets playing with some toys and saying some things about them and that then they would be asked some questions about the puppets' statements. Next, the experimenter familiarized children with the use of a 6-point Likert-type scale in the form of faces with smiles and frowns of varying degrees. Children were given practice at judging the acceptability of different actions (e.g., helping individuals clean up their toys vs. pushing them) on a scale from 1 = *really not okay* to 6 = *really okay.*

After this familiarization, children saw two trials each of three trial types: *full verification, insufficient verification,* and *no verification.* On each trial (counterbalanced for order of within-participant conditions), children saw a storybook page in which a puppet encountered four unusual-looking containers and expressed that he (the puppet) did not know anything about them. On full verification trials, children saw that the puppet looked inside each of the four containers. On insufficient verification trials, they saw that the puppet looked in only one of the four containers and deliberately chose not to look in the other three containers. On no verification trials, they saw that the puppet deliberately chose not to look in any of the containers. On all trials, the puppet then made a claim about the contents of all four containers (e.g., "All of these have marbles inside!"). Thus, the claims were based on sufficient, insufficient, or no evidence.

On each trial, children were asked three questions: (a) whether it is okay or not okay for the puppet to have said that all of the containers had particular contents, (b) how okay or not okay that claim was using the 6-point Likert-type scale, and (c) why they judged the statement to be okay or not okay. It is important to note that in asking children to judge the acceptability of these claims, we were not asking children to make trait-based inferences about who is more reliable or trustworthy. Rather, we were asking them to evaluate whether making that assertive speech act was acceptable on the basis of the puppet's evidentiary practices.

Results

Acceptability ratings

The residuals of children's acceptability ratings were normally distributed and met the requirements for parametric statistics. Children's mean acceptability ratings across the three trial types (see Fig. 1) were analyzed using a repeated-measures analysis of variance (ANOVA), with condition and trial (first or second) as within-participant factors and age group (preschool or elementary school) and order as between-participant factors and with post hoc Tukey's HSD (honestly significant difference) tests to guard against issues of multiple comparison. This analysis revealed a significant effect of condition, $F(2, 84) = 26.50$, $p < .001$, partial $\eta^2 = .387$, and a main effect of age group, $F(1, 42) = 8.39$, $p = .006$, partial $\eta^2 = .166$. There were no other significant main or interaction effects. Children evaluated the puppets' statements on the full verification trials ($M = 4.51$, $SD = 1.49$) as significantly more acceptable than on either insufficient verification trials ($M = 3.16$, $SD = 1.23$), $p < .001$, or no verification trials ($M = 2.77$, $SD = 1.52$), $p < .001$. There was no statistically significant difference between the insufficient verification and no verification trials ($p = .259$).

To investigate this developmental difference, we analyzed each age group separately using the same analysis as described above. For the preschool-aged children, this analysis revealed a significant effect of condition, $F(2, 42) = 10.60$, $p < .001$, partial $\eta^2 = .347$. Children rated the puppets' statements on the full verification trials ($M = 4.25$, $SD = 1.77$) as significantly more acceptable than on either insufficient verification trials ($M = 2.79$, $SD = 1.19$), $p = .010$, or no verification trials ($M = 2.25$, $SD = 1.29$), $p < .001$. There was no significant difference between the insufficient verification and no verification trials ($p = .335$).

For the elementary school-aged children, the analysis revealed the same pattern. There was a significant effect of condition, $F(2, 42) = 16.10$, $p < .001$, partial $\eta^2 = .434$. Children rated the puppets' statements on the full verification trials ($M = 4.81$, $SD = 1.11$) as significantly more acceptable than on either the insufficient verification trials ($M = 3.52$, $SD = 1.18$), $p < .001$, or no verification trials ($M = 3.29$, $SD = 1.59$), $p < .001$.There was no significant difference between the insufficient verification and no verification trials ($p = .709$).

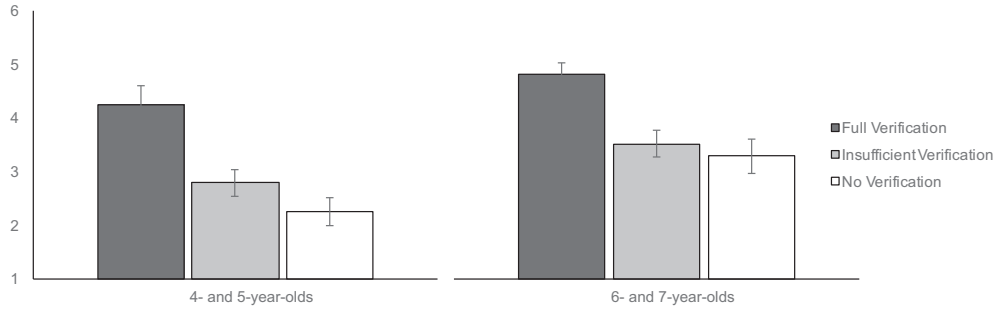

Fig. 1. Mean acceptability ratings by condition in Experiment 1. Error bars represent ± 1 standard error of the mean.

Explanations

To assess children's explicit understanding of verification, we coded children's explanations (their answers as to why they rated each statement as okay or not okay) according to whether or not they explicitly referenced epistemic states or actions relevant to verification. An explanation was coded as a verification response if it explicitly referenced the puppet checking the boxes (e.g., "He didn't check," "He checked what was in them"), opening them (e.g., "He didn't open it", "He opened the box"), looking in them (e.g., "Because he didn't look," "Because he looked inside"), or knowing about their contents (e.g., "Because he doesn't know that," "Because he knows what's in them"). Any other explanation, including epistemically relevant responses such as guessing and thinking that nevertheless do not reference verification per se, were coded as non-verification responses. Two blind coders provided independent ratings on the full set of explanations using this coding scheme. There were no disagreements between the two coders.

On average, children gave verification explanations on 52.2% of trials ($M = 3.13$, $SD = 1.84$). To test the relation between verification and justification, we next analyzed whether the number of trials in which children gave verification explanations varied by condition. A repeated-measures ANOVA with condition as a within-participant factor and age group and order as between-participant factors revealed a main effect of condition, $F(2, 84) = 22.60$, $p < .001$, partial $\eta^2 = .350$, and a marginal main effect of age group, $F(1, 42) = 3.92$, $p = .054$, partial $\eta^2 = .085$, with elementary school-age children being slightly more likely to give verification explanations. Overall, children did so more in the insufficient verification ($p = .065$) and no verification ($p < .001$) conditions than in the full verification condition and did so more in the no verification condition than in the insufficient verification condition ($p < .001$).

Relation between acceptability ratings and explanations

We next asked whether children's acceptability ratings were related to their explicit explanations of those ratings. To do so, we calculated a difference score between children's mean acceptability ratings on the full verification trials and their mean acceptability ratings across both types of unverified trial types (both insufficient verification and no verification trials). Next, we analyzed whether this difference score was correlated with the extent to which children gave verification explanations across the three within-participant conditions. We found that this difference score was highly correlated with the number of trials in which children explicitly referenced epistemic reasons in their explanations in all three conditions (full verification: $r = .352$, $p = .014$; insufficient verification: $r = .354$, $p = .013$; no verification: $r = .304$, $p = .036$). Thus, children who gave verification explanations on more trials also made a stronger differentiation between verified and unverified claims.

Discussion

Taken together, the results of Experiment 1 suggest that by preschool age children are aware of the importance of verification in making factual claims and that claims that have not been verified are less acceptable than claims that have been verified. Furthermore, this understanding, although present even in the younger group of children, develops with age. In addition, the fact that children at both ages rated completely unverified and insufficiently verified claims similarly suggests that at these ages they have some understanding of what kind or extent of verification is needed to credibly make certain types of claims. Finally, children's explanations of their evaluations suggest that their success in differentiating verified and unverified claims is based on an explicit understanding of the process of verification.

It is important to point out that even in the insufficient verification and no verification conditions, children nevertheless rated the puppets' claims in approximately the middle of the scale, somewhere between *a little not okay* and *a little okay*. Thus, although they distinguished between verified and unverified claims, they did not seem to view unverified claims as completely unacceptable—simply less acceptable than verified ones. This suggests that although children are sensitive to verification in their evaluation of claims, they might not be particularly skeptical of unverified claims. This may be because children have a documented tendency to trust others' claims (see Jaswal, Croft, Setia, &

Cole, 2010; Koenig & Jaswal, 2011) as well as because the claims in this case were plausible and potentially true; they simply had not been verified.

Furthermore, older children seemed to rate the unverified claims as somewhat more acceptable than did younger children. Why might this be? Although we do not have direct evidence to speak to this issue, there are several possibilities. First, older children may understand more about how knowledge acquisition works and that there are cases in which someone has access to information that is not immediately apparent. Thus, they might be more hesitant to ascribe deliberate ignorance or negligence, especially in cases like this where the claims are both plausible and potentially true. Younger children may well be less nuanced in their reasoning. Second, older children may have been more likely to compare the current unverified claims with the less acceptable item used in familiarizing them with the scale (pushing someone) and reasoned that failing to verify a plausible claim paled in comparison with that kind of moral violation. Relatedly, as we discuss later, how children interpreted okay and not okay is an open question that future research will need to address; if children's interpretation differed with age, this may well have affected their ratings.

Overall, the results of this experiment demonstrate that, beyond simply using their epistemic reasoning skills to choose who is more likely to be a knowledgeable informant, children are able to evaluate factual claims based on the evidentiary process supporting them. Furthermore, to some extent, they are able to explain why those claims are or are not acceptable. This is an important step forward for our understanding of epistemic reasoning during the early years of life. This finding suggests that children can go beyond mere informant reliability and can reason about the underlying empirical process by which valid claims are made and supported by evidence.

Experiment 2

Experiment 2 served two purposes. First, we aimed to replicate the findings of Experiment 1 with a new sample of children. Second, we aimed to expand the developmental investigation, asking whether preschoolers across a broader age range can reason about verification to assess the legitimacy of differentially verified claims. Given the prior lack of differentiation between full and insufficient verification, and concerns about task demands with younger children, we simplified the task and developed more vivid scenarios in which to present individuals' claims, as described below.

Method

Participants

In Experiment 2, we tested an additional 72 preschool-aged children in three age groups (3-year-olds: mean age = 40.8 months, range = 36–46, 12 boys; 4-year-olds: mean age = 53.7 months, range = 48–59, 12 boys; and 5-year-olds: mean age = 66.2 months, range = 60–71, 12 boys). Because of concerns about distractions at the museum, especially for the youngest children, this sample was recruited solely into the lab at a large university in a nearby town and tested in a private research room. As in Experiment 1, the sample was largely upper-middle class, but children represented a range of ethnic backgrounds, with the sample being 62% White/Caucasian, 15% Black/African American, and 23% from other or mixed ethnic backgrounds.

Procedure

The materials and methodology were similar to those in Experiment 1, with several key changes. On each of three test trials, children saw two characters (Raccoon and Bear), who were said to be friends, in a real-life setting (outside their homes, in their classroom at school, or outside after school). On each trial, each of the characters discovered an unfamiliar container and wondered what was inside it. One character looked inside the container and made a factual claim about what was inside, whereas the other character chose not to look inside the container and nevertheless made a factual claim about what was inside. The exact nature of the claims varied, but each followed the same logic; at home the characters received surprise packages and made claims about what was in the package (Legos vs. blocks), at school the characters found pencil boxes and made claims about what color

the pencils were (blue vs. green), and after school the characters got gift bags with stickers and claimed what kind of stickers were inside (square stickers vs. circle stickers). On each trial, each individual's evidentiary behavior (checking or not checking) and claims were presented sequentially (counterbalanced between participants both for order and for which character made verified or unverified claims). After each individual's claim, children were asked the same acceptability questions as in Experiment 1.

Results

Acceptability ratings

The residuals of children's acceptability ratings were normally distributed and met the requirements for parametric statistics. As in Experiment 1, children's mean acceptability ratings (1 = *really not okay* to 6 = *really okay*) across the three trials were analyzed using a repeated-measures ANOVA, with condition and trial as within-participant factors and age group and order as between-participant factors, with a post hoc Tukey HSD test. This analysis revealed a significant main effect of condition, $F(1, 60) = 18.20$, $p < .001$, partial $\eta^2 = .233$ (Fig. 2). There were no other significant main or interactive effects. Overall, children rated the verified claims ($M = 4.44$, $SD = 1.53$) significantly higher than the unverified ones ($M = 3.55$, $SD = 1.59$), $p < .001$.

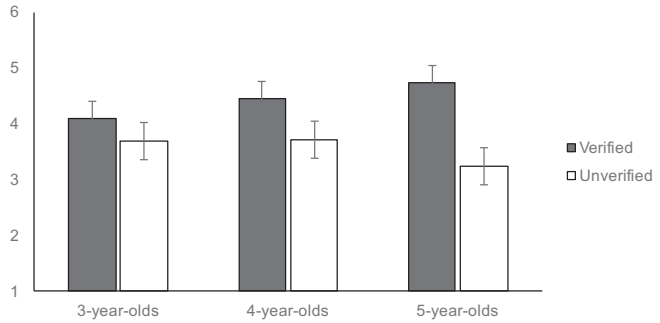

Fig. 2. Mean acceptability ratings by condition in Experiment 2. Error bars represent ± 1 standard error of the mean.

Explanations

To assess children's explicit understanding of verification, we coded children's explanations using the same coding scheme as in Experiment 1. Two blind coders provided independent ratings on the full set of explanations using this coding scheme, and there were no disagreements between the two coders.

We next analyzed whether the number of trials on which children gave verification explanations varied by condition and age group. A repeated-measures ANOVA with condition as a within-participant factor and age group and order as between-participant factors revealed only a main effect of age group, $F(2, 60) = 11.60$, $p < .001$, partial $\eta^2 = .279$. There were no other significant main or interactive effects. The 5-year-olds gave verification explanations significantly more overall ($M = 2.58$, $SD = 2.39$) than either the 4-year-olds ($M = 1.00$, $SD = 1.60$), $p = .006$, or the 3-year-olds ($M = 0.292$, $SD = 0.624$), $p < .001$.

Relation between acceptability ratings and explanations

Echoing the findings from Experiment 1, we found that the difference score between children's mean acceptability ratings on the verification and no verification trials was highly correlated with the number of trials on which children gave verification reasons in their explanations on both trial types (verification: $r = .437$, $p < .001$; no verification: $r = .302$, $p = .010$). Thus, children who gave verification reasons on more trials also made a stronger differentiation between the verified and unverified claims.

Discussion

The results of Experiment 2 suggest that even children as young as 3 years are aware of the importance of verification in making claims and that claims that have not been sufficiently verified are less acceptable than claims that have been verified. Furthermore, the extent to which children differentiated verified and unverified claims in their acceptability ratings was correlated with the extent to which they explicitly referenced verification in their explanations, and the ability to do so changed with age.

General discussion

We live in a society increasingly inundated with information, where we continually need to scrutinize others' claims in order to discern accurate and robust knowledge. Even several years ago it was estimated that more than 90% of the information ever produced was generated during the preceding 2 years (SINTEF, 2013) and the amount of information in the digital world is both constantly growing and increasingly accessible. Growing up in a digital world where it would be impossible to gauge the accuracy of every piece of information, the critical thinking skills necessary to assess the validity of a source is crucial for children to learn. This emphasis is reflected in recent educational standards. A principal goal of both the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the Next Generation Science Standards (GSS Lead States, 2013) is to foster children's scientific "habits of mind," the ways in which children approach inquiry and reasoning about evidence in order to support scientific thinking. As early as kindergarten, this type of empirical reasoning is emphasized in these standards, and empirical reasoning plays a central role in readiness for STEM learning. Experts from education and policy emphasize that teaching STEM skills early in childhood is fundamentally important (Bowman, Donovan, & Burns, 2001; Duschl, Schweingruber, & Shouse, 2007; Eshach & Fried, 2005; Gelman & Brenneman, 2004). Indeed, recent findings indicate that STEM knowledge at kindergarten entry is the best predictor of achievement in science during the elementary school years (Morgan, Farkas, Hillemeier, & Maczuga, 2016). Although an important aspect of science education is teaching a core set of scientific facts in specific content domains, understanding the *process* by which scientists engage in critical evaluation of empirical evidence is arguably the lynchpin of readiness for science. The National Research Council (2000) recommended that science education during the early years not only teach key scientific concepts but also facilitate the development of scientific inquiry skills.

The current research shows that 3- to 7-year-old children, even prior to formal education that might train them in this kind of reasoning, understand that claims that rest on a sufficient body of evidence are more acceptable than insufficiently verified ones. Furthermore, the current experiments show that there is a direct relationship between children's explicit understanding of the importance of verification, as indexed by their spontaneous reference to epistemic reasons when asked to explain their judgments, and the extent to which they differentiate between verified and unverified claims.

This work sets the stage for a number of follow-up questions further examining children's developing understanding of verification and its importance. First, in the current experiments, we found a correlation between explicitly mentioning verification and differentiation between verified and unverified claims in their acceptability ratings. However, the current work does not tell us the exact nature of this relationship. Is an understanding that verified claims are preferable to unverified ones predicated on an explicit understanding of why this is true, or is there an age at which children have an implicit or intuitive preference for verified claims over unverified ones without being able to explain why? Furthermore, to the extent that the ability to explain evaluations changes with age, what mechanisms underlie this developmental change? Future research will need to delve into possible mechanisms, including changes in social, cognitive, linguistic, and executive function capacities, each of which may plausibly play key roles in supporting children's empirical reasoning.

Second, to what extent are children reasoning about verification along social and normative dimensions? In the current work, children judged verified claims to be more okay than unverified ones. But the term *okay* is ambiguous and could be understood to mean "normal" (i.e., people typically attempt

to verify their claims before making them), "valid given the evidence," or "normatively acceptable given our conventions around evidence and claim making." Prior research has shown that children are on the lookout for norm-relevant actions, ascribing normativity widely on the basis of intention without explicit normative cues (Butler, Schmidt, Bürgel, & Tomasello, 2015; Schmidt, Butler, Heinz, & Tomasello, 2016; Schmidt, Rakoczy, & Tomasello, 2011) and doing so on the basis of simple regularities (Roberts, Gelman, and Ho, 2016; Roberts, Guo, Ho, & Gelman, 2018; Tworek & Cimpian, 2016). Thus, it seems entirely plausible that children interpret *okay* in a normative light. Nevertheless, future research is needed to directly test this possibility, for example, by examining whether children might spontaneously object to or protest against unverified claims.

Third, it is unknown to what extent children understand the concept of verification per se. In the current research, children were able to understand that an individual who failed to look inside a container could not (reliably) make valid claims about its contents. And children show at least some understanding that different knowledge is needed to make claims about an object's identity versus its name (Brosseau-Liard & Birch, 2011). But in developing the connection between these findings and the development of critical thinking and skepticism more broadly, it is necessary to drill down further into the richness of children's understanding. To what extent do children understand that different types of claims require different types of evidence and that what serves as sufficient verification for one type of claim may be insufficient for another?

Finally, the current work focused on the claims themselves rather than on the individuals making the claims. There is a long literature on children's evaluation of informants' accuracy and trustworthiness (see Harris et al., 2018). Thus, it would be important to ask whether children's evaluation of verification would carry over to their inferences about who to learn from or who to trust. That is to say, how do children integrate top-down processes (using trait-based cues to make judgments about the reliability of claims) with bottom-up processes (using evidentiary behaviors to make judgments about an individual's traits, e.g., reliability)? In addition, recent work has shown that children's preference for informants also depends in some cases on their social identity (e.g., racial, ethnic, or other social group membership; see Corriveau et al., 2013; Kinzler, Corriveau, & Harris, 2011). Might children be more or less accepting of unverified claims made by or about members of different social groups? Might children be more accepting of unverified claims made in a pedagogical manner, which has been shown to lead to inferences that are more resistant to counterevidence even in very young children (Butler & Markman, 2016; Butler & Tomasello, 2016; Hernik & Csibra, 2015), or by a knowledgeable and helpful teacher (Gweon & Asaba, 2018; Gweon, Pelton, Konopka, & Schulz, 2014)? Or might children privilege past accuracy or expertise over current empirical practices? More broadly, how do children weigh epistemic and cultural or social cues, especially when they conflict?

This research is especially important as we aim to raise the next generation of critical consumers of evidence, particularly in the STEM fields. By moving away from specific causal relations or scientific domains, and by focusing not on the individuals making claims but rather on the process by which those claims are supported by evidence, the current work addressed this issue directly by focusing directly on children's evaluations of the empirical process itself. The current research lays the groundwork for an important line of research investigating the roots and developmental course of this empirical reasoning process during early childhood and its relation to critical thinking and learning across a broad range of domains in both formal and informal educational settings. By investigating children's understanding that claims must be verified in order to be accepted at the age when they are just entering school, this research has the potential to help pinpoint the developmental timing and trajectory of understanding and, thus, generate potential targets for intervening to foster children's critical thinking and empirical reasoning skills.

Acknowledgments

This article was supported in part by a grant from the Spencer Foundation to the first author and by a grant from the Elite Network of Bavaria, an initiative of the Bavarian State Ministry for Education, Science, and the Arts, to the second author. We thank Karen Levush for assistance with coding.

References

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to learn*. Washington, DC: National Research Council.

Brosseau-Liard, P. E., & Birch, S. A. J. (2011). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development, 82*, 1788–1796.

Butler, L. P., & Markman, E. M. (2016). Navigating pedagogy: Children's developing capacities for learning from pedagogical interactions. *Cognitive Development, 38*, 27–35.

Butler, L. P., Schmidt, M. F. H., Bürgel, J., & Tomasello, M. (2015). Young children use pedagogical cues to modulate the strength of normative inferences. *British Journal of Developmental Psychology, 33*, 476–488.

Butler, L. P., & Tomasello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology, 145*, 64–78.

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language, 19*, 360–379.

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

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in Grades K–8*. Washington, DC: National Academies Press.

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

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology, 14*, 315–336.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60*, 115–140.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly, 19*, 150–158.

GSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Gweon, H., & Asaba, M. (2018). Order matters: Children's evaluation of underinformative teachers depends on context. *Child Development, 89*, e278–e292.

Gweon, H., Pelton, H., Konopka, J. A., & Schulz, L. E. (2014). Sins of omission: Children selectively explore when teachers are under-informative. *Cognition, 132*, 335–341.

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

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

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology, 69*, 251–273.

Hernik, M., & Csibra, G. (2015). Infants learn enduring functions of novel tools from action demonstrations. *Journal of Experimental Child Psychology, 130*, 176–192.

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

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

Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development, 83*, 1051–1063.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15*, 694–698.

Koenig, M. A., Cole, C. A., Meyer, M., Ridge, K. E., Kushnir, T., & Gelman, S. A. (2015). Reasoning about knowledge: Children's evaluations of generality and verifiability. *Cognitive Psychology, 83*, 22–39.

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

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects? *Child Development, 82*, 1634–1647.

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

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*, 18–35.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards*. Washington, DC: Authors.

National Research Council (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning science and mathematics program improvement*. Washington, DC: National Academy Press.

Nurmsoo, E., & Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well- or poorly-informed: When past errors can be excused. *Child Development, 80*, 23–27.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*, 116–129.

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2016). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science, 41*, 576–600.

Roberts, S. O., Guo, C., Ho, A. K., & Gelman, S. A. (2018). Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally: Evidence from China. *Journal of Experimental Child Psychology, 165*, 148–160.

Robinson, E. J., Butterfill, S. A., & Nurmsoo, E. (2011). Gaining knowledge via other minds: Children's flexible trust in others as sources of information. *British Journal of Developmental Psychology, 29*, 961–980.

Schmidt, M. F. H., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young children see a single action and infer a social norm: Promiscuous normativity in 3-year-olds. *Psychological Science, 27*, 1360–1370.

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2011). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition, 124*, 325–333.

SINTEF (2013, May 22). Big data, for better or worse: 90% of world's data generated over last two years. ScienceDaily. Retrieved July 18, 2017, from http://www.sciencedaily.com/releases/2013/05/130522085217.htm.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*, 359–393.

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science, 27*, 1109–1122.