Children flexibly adapt their evidentiary standards to their informational environments

Evan Orticio, Martin Meyer, and Celeste Kidd
Department of Psychology, University of California, Berkeley, USA

Abstract

How do children decide when to fact-check a claim? In two studies and a simulation, we show that children use environmental cues about informational reliability to rationally adapt their information seeking to enable the efficient detection of misinformation. In Study 1, children exposed to an environment containing some misinformation (as opposed to all true information) sampled more evidence before verifying a test claim in a novel domain. In Study 2, children showed a graded sensitivity to environmental reliability: information seeking increased linearly with the proportion of false statements heard during exposure. Additionally, these statements were presented as online search results, which demonstrated that children make sophisticated inferences about their informational environments, beyond speaker-specific cues, to adjust their skepticism toward new information. These results further emphasize the importance of considering the environmental context of learning when developing interventions to promote healthy skepticism and lifelong learning.

Keywords: information seeking; skepticism; informational environment; misinformation

Introduction

Modern informational environments like the internet are a double-edged sword. Online platforms and communities offer historically unprecedented opportunities for knowledge exchange. However, they also leave users exposed to unprecedented amounts of misinformation. Exposure to misinformation alone can lead to the adoption of false beliefs, even when the learner is aware of this bias (Fazio et al., 2015). Adopted misinformed beliefs can persist for months after exposure (Brown & Nix, 1996). Further, children possess characteristics that could leave them more susceptible to misinformation than adults; they are less media literate and possess less domain-relevant knowledge to evaluate novel claims.

Children's vulnerability to misinformation is a threat given how early they immerse themselves in online environments. By age 7-9, a third of American children are already on at least one social media platform (Mott Poll, 2021). The media habits that children form at this early age likely carry into adolescence and beyond. A majority of American teens report getting their news from social media or YouTube rather than directly from news organizations (Common Sense Media, 2019), suggesting that their media diets are filled with potentially non-credible sources. How do we best prepare children to navigate complex informational environments like social media and the internet?

One approach to this problem has been to heavily control children's informational diets and to shield them from misinformation in the first place. Moderated platforms and search engines targeted toward children are on the rise. A leader in this domain, YouTube Kids, currently offers a small selection of human-curated content and a suite of fine-grained parental controls (Rodriguez, 2018). While platforms like these can be a helpful start, full content moderation is unrealistic. As an example, YouTube Kids has fallen under controversy several times in recent years following discoveries of disturbing videos slipping through filters (Maheshwari, 2017). Even at its best, content moderation is resource-intensive, relies on subjective decisions about what is considered age-appropriate content, and is likely to be deployed inequitably (e.g., by prioritizing English language content).

We suggest a second potential drawback of this approach. Sanitized informational environments could pose an unintended negative consequence for children by inadvertently promoting unjustifiably trusting media consumption habits. In addition, overly moderated content could rob children of opportunities necessary to develop the critical thinking skills they will need to navigate more complex environments later in life.

We propose an alternate approach to preparing children for misinformation throughout their development. Our proposed approach embraces the complexities of modern informational environments and allows children to build upon their natural truth-seeking competencies. While kids may struggle to navigate completely unmoderated contexts safely, an environment that exposes kids to some falsity may prepare them to detect more subtle misinformation by providing the opportunity to develop their critical thinking skills. Here, we test whether controlled but imperfect informational environments may serve as useful scaffolding.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Figure 1: Test phase, identical in Studies 1 and 2. After checking a zorpie (e.g., bottom row, third from left), children could choose to accept the statement (green button), reject the statement (red button), or check another zorpie before deciding (magnifying glass).

Study 1

Study 1 asks whether children use the prior reliability of information in their environment to shape their standards of evidence for a novel claim. Do children increase their evidentiary standards for a claim selectively after exposure to misinformation? To test this, children were randomly assigned to judge the veracity of a set of animal facts that were either all true (Reliable condition) or partially false (Unreliable condition). Following this, children judged a novel claim about aliens, and were given the opportunity to freely sample evidence about the claim before making their final decision. We hypothesized that children would sample more evidence before trusting the claim in the Unreliable condition.

Method

Participants Sixty 4- to 6-year-old children (M_{age} = 5.51, SD = 0.89, 47% White) were recruited from parks in the California Bay Area. Three additional children were excluded from analysis because they had watched another child participate or were too distracted to complete the study.

Procedure Children used a touchscreen computer to play a game created in PsychoPy. In an exposure phase, the experimenter asked children to determine whether a set of statements about animals in an e-book were right or wrong. On each of 12 exposure trials, the tablet displayed a statement (e.g., "Zebras have black and white stripes") and an accompanying picture. Children first tapped a button to hear an audio recording of the statement, and then indicated whether the fact was right (by tapping a green button) or wrong (by tapping a red button). The facts varied by condition (between-subjects, pseudo-randomly assigned). In the Reliable condition, all 12 animal statements were true. In the Unreliable condition, four of the 12 animal statements were clearly false (e.g., "Zebras have red and green stripes"). Pictures were identical across conditions, so children could judge the statements using real-world knowledge or the pictures alone.

In the subsequent test phase (see Figure 1), children moved on to a second chapter of the e-book which was about a novel alien species called "zorpies". They were asked to evaluate a new statement about zorpies: "All zorpies have exactly three eyes under their glasses." The screen displayed the fact alongside 20 zorpies wearing opaque sunglasses. After tapping a button to hear the fact, children were told that they could tap any zorpie to remove its glasses and reveal its eyes. All zorpies were identical and had three eyes, so any evidence the child sampled supported the test statement. Once the child tapped a zorpie, they had to decide to tap a button to accept the statement as true, reject the statement as false, or to check another zorpie first. This procedure repeated such that children could check as many zorpies as they wished (from 1 to all 20) before indicating whether the fact was right or wrong and completing the study. The task was designed to produce different information seeking behavior depending on one's level of skepticism toward the claim. A fully trusting learner might see that all the zorpies are identical and be satisfied after checking only one, while a highly skeptical learner might feel the need to check all 20 zorpies because the statement refers to "*all* zorpies".

Results and Discussion

Manipulation check Children reliably discerned between true and false statements in the exposure phase of the experiment. Children's accuracy in evaluating statements as true or false were significantly above chance in both the reliable (M = 9.40 of 10 correct, $p < .001$) and unreliable conditions (M = 8.43 of 10 correct, $p < .001$), indicating that we successfully manipulated the reliability of participants' informational environments. Nine of the 60 participants failed to achieve 80% accuracy, but their exclusion does not affect any results, so we maintain their data for all future analyses. Additionally, all but three children (95%) correctly judged the test claim to be true, suggesting that children were generally tracking the evidence appropriately. Of the three participants who rejected the test claim, two were in the unreliable condition.

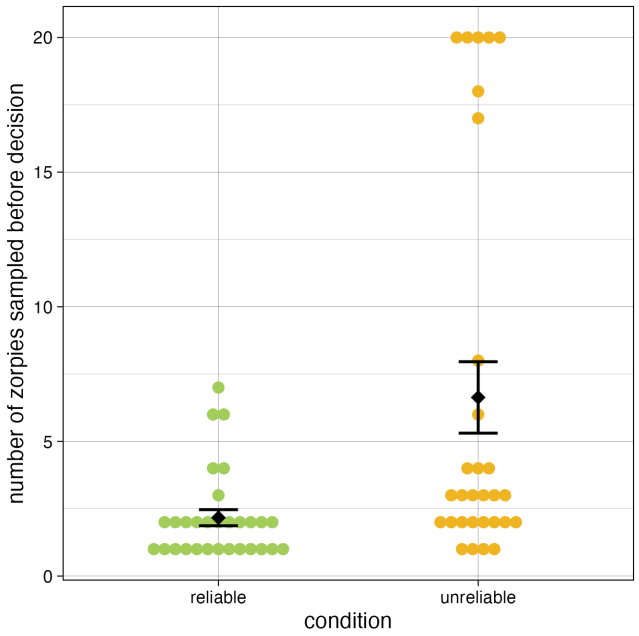

Figure 2: Children sampled more evidence in the unreliable environment. Dots are individual data points, diamonds are condition means, and error bars represent one SEM. The effect of condition remains robust after winsorization, ensuring that the highest sample values do not drive the effect.

Children seek more evidence in unreliable environments Children increase their standards of evidence for new claims in an environment containing some misinformation. Figure 2 shows the number of zorpies children sampled before deciding to accept or reject the test claim by condition. On average, children in the Unreliable condition sampled more evidence than those in the Reliable condition (M = 6.63 vs. 2.17 zorpies, Wilcoxon W = 233, $p <$ .001). In the environment containing some misinformation, children sought out significantly more information before deciding whether to accept the test claim. In the Unreliable condition, a number of children even opted for an exhaustive or near-exhaustive sampling strategy, checking up to 20 zorpies in a row even though all the prior evidence was identical. Children were thus able to leverage the prior quality of their informational environment in a known domain (animal facts) in order to adapt their skepticism and subsequent information search about a novel claim about which they had no prior knowledge.

The distribution of sampling behavior in the unreliable condition was bimodal, so we also winsorized the data such that the maximum value of zorpies sampled was 8 (the maximum of the other mode). The fact that an exhaustive sampling strategy leads children to sample exactly 20 zorpies is the result of a design choice, so replacing extreme values with the maximum value of the other mode provides a more stringent and design-neutral test of our hypothesis. The effect remained robust after winsorization (W = 233, $p < .001$), suggesting that it was not driven by the subset of exhaustive samplers in the unreliable condition.

Finally, we tested whether children's sensitivity to the reliability of their informational environments changes with age. We used the *MASS* package in R to run a robust linear regression using standardized age and condition to predict the number of zorpies sampled. This analysis replicated the main effect of condition (β = 1.41, p = .007), but revealed no main effect of age and no interaction (p's > .05). There were no changes in how children responded to the reliability of information from ages 4 through 6 in our sample.

Study 2

The informational reliability of an environment is not all-or-nothing. Do children appreciate nuances in the reliability of their environments, and adapt their level of skepticism accordingly? To address this question, Study 2 introduced five between-subjects conditions of varying reliability, ranging from 0% to 80% false statements in the exposure phase. Additionally, children likely conceptualized the environment in Study 1 as a single cohesive source. The task was framed as an e-book, and children heard all statements in audio recordings using the same voice. Can children still make smart inferences about claims based on a more complex environment composed of many distinct sources? In Study 2, we presented statements as individual search results, read in distinct voices, to test whether children make more abstract generalizations about their environment to adapt their information seeking.

Method

Participants Sixty-two 4- to 7-year-old children (M_{age} = 5.88, SD = 1.06, 52% White) were recruited from parks in the California Bay Area. Four additional participants were excluded from analysis because they had watched another child participate. None of the Study 2 participants had completed Study 1 previously.

Procedure The procedure was identical to Study 1 aside from two main changes. First, we created five between-subjects

conditions such that 0%, 20%, 40%, 60%, or 80% of the 10 exposure trials were false statements. Second, the activity was reframed so that the statements appeared to originate from distinct search engine results. The experimenter typed "Animal facts" into a search bar to generate a simulated results page in the exposure phase. The experimenter tapped on a search result to begin a trial. On each trial, the style of the picture and the voice of the audio recording were different. The test phase followed. In the test phase, the experimenter input "Alien facts" into the search bar and tapped a result to display the page of zorpies. The audio recording on the zorpie test trial featured another distinct voice.

Results and Discussion

Manipulation check Children reliably discerned between true and false statements in the exposure phase of the experiment in Study 2. Accuracy in the exposure phase was significantly above chance ($M = 9.24$ of 10 correct, $p < .001$). Four of the 62 participants failed to achieve 80% accuracy, but their exclusion does not affect any results, so we maintain their data for all future analyses. Additionally, all but four children (93.5%) correctly judged the test claim to be true. The children who rejected the test claim were in the two most unreliable conditions (three in the 80% false condition, one in the 60% false condition).

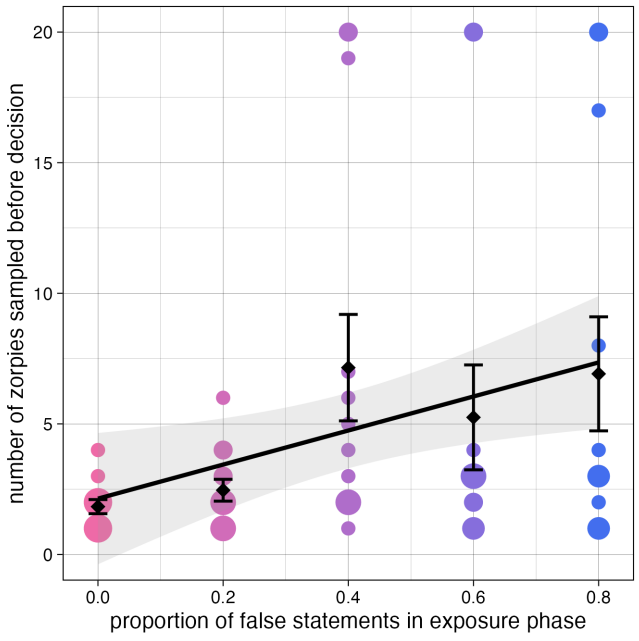

Figure 3: Children sampled more evidence as the reliability of their environments decreased. Amount of evidence sampled (out of a possible 20 zorpies) vs. proportion of false statements in exposure phase in Experiment 2. The size of the dot represents the number of data points. Diamonds are conditional means and error bars represent one SEM. Line represents the fit from linear regression with a 95% confidence interval.

Graded sensitivity to environmental reliability Figure 3 shows the number of zorpies children sampled before deciding to accept or reject the test claim as a function of the proportion of false statements encountered in the exposure phase. A linear regression revealed that the proportion of false statements in the exposure phase was a significant linear predictor of the number of zorpies sampled ($\beta = 6.52$, $p = .014$). Skepticism increased linearly with increases in the number of false statements in the exposure phase, manifesting in more extensive information search in the test phase. Children are thus able to make sophisticated, graded judgments about the reliability of their current informational environments, and use that to guide future learning. Note that this sensitivity was observed in a simulated search engine environment composed of distinct sources–each statement was heard from a different voice. This suggests that children went beyond speaker-based heuristics and tracked the cumulative quality of information throughout the exposure phase.

Because the data was bimodal, we replicated this result with a robust linear regression, which is less sensitive to influential observations. The proportion of false statements remained a significant linear predictor ($\beta = 1.95$, $p = .049$). These findings suggest that, at a group level, children linearly scaled up their evidentiary standards as the reliability of their informational environments decreased.

While we observed a linear relationship between environmental reliability and information sampling in the present data (as predicted), it is unclear whether this would generalize across all tasks. In our task, the information to be gained by additional sampling was maximally transparent. The outcome was binary and directly related to the claim in question (the next zorpie is three-eyed or not), and the full space of available evidence was clearly delineated. Other studies with low-risk exploration have also found linear associations between low certainty and information seeking in children (Coughlin et al., 2014) and adults (Desender et al., 2018).

On the other hand, some evidence suggests environments characterized by variable expected information gain induce a U-shaped relationship between curiosity and information seeking (Baranes, Oudeyer, & Gottlieb, 2014; Wang & Bonawitz, 2019). This pattern of results is consistent with a dual-process account of metacognition, in which information seeking is guided not only by certainty, but also by an appraisal of the potential information gain afforded by the environment (Goupil & Proust, 2023; Baer & Kidd, 2022). Speculatively, the bimodal distribution of sampling strategies even in the most unreliable conditions of Study 2 may represent two distinct interpretations of the environment. Some of the children who checked only a few zorpies in highly unreliable environments may have been highly skeptical, but doubted that the available evidence would provide accurate information in the first place. The linear effect we observe may therefore be the combination of two patterns of responses: an even stronger linear effect dampened by a group exhibiting a U-shaped pattern.

Sensitivity moderately increases with age Does children's skepticism become more finely attuned to the reliability of their informational environment as they age? We ran a robust linear regression using standardized age and the standardized proportion of false statements in exposure to predict the number of zorpies sampled. First, we replicated the main effect of environmental reliability (i.e., proportion of false statements in exposure phase, $\beta = 1.29$, $p < .001$), demonstrating that this effect is robust using an analysis that is less sensitive to influential observations. In addition, this analysis revealed a significant main effect of age ($\beta = 1.49$, $p < .001$) and a significant interaction ($\beta = 1.22$, $p < .001$). The main effect of age suggests that older children sought out more evidence than younger children overall. The reliability by age interaction suggests older children were more sensitive than younger children to variation in environmental reliability. Older children in our sample, and particularly the 7-year-olds, were more likely than younger children to sample a high number of zorpies when their environments contained a high proportion of false information in the past.

Simulation

In Studies 1 and 2, the test claims were true. Yet, the selective skepticism that children exhibit in these studies is theoretically adaptive because increased information sampling facilitates the discovery of counterevidence. If the test claim was actually false, what kind of environment would best prepare children to discover that? We ran a simulation to determine whether experience learning in an unreliable environment enables children to identify misinformation more easily. In each of 400 simulation runs, we randomly sampled a proportion of the available zorpies to serve as hypothetical counterevidence to the test claim (i.e., zorpies *without* three eyes). Then, using participants' real sampling behavior from Study 1, we calculated the percentage of participants in each condition who revealed one or more of the counterevidence zorpies before making a decision about the test claim. Thus, each run of the simulation represents a hypothetical experiment with potential counterevidence randomly distributed among the 20 zorpies. We simulated four different proportions of counterevidence in the sample space with 100 simulation runs each, such that either 10%, 20%, 30%, or 40% of the available zorpies provided evidence against the test claim. The results of the simulation are in Figure 4.

To test whether children in the unreliable condition were more likely to discover counterevidence during sampling, we used the *betareg* package in R to run a beta regression using experimental condition (reliable vs. unreliable) to predict the percentage of Study 1 participants who would have discovered one or more pieces of counterevidence during sampling. This analysis revealed a main effect of condition ($\beta = 0.86$, $p < .001$), and the effect holds after controlling for the simulated amount of counterevidence ($\beta = 0.93$, $p < .001$). These results support the commonsense conclusion that children in the unreliable condition, who sampled more evidence bearing on the test claim, would have been more likely to discover counterevidence had it been available. Unreliable learning environments elicit increased skepticism and thus enable children to debunk misinformation more readily.

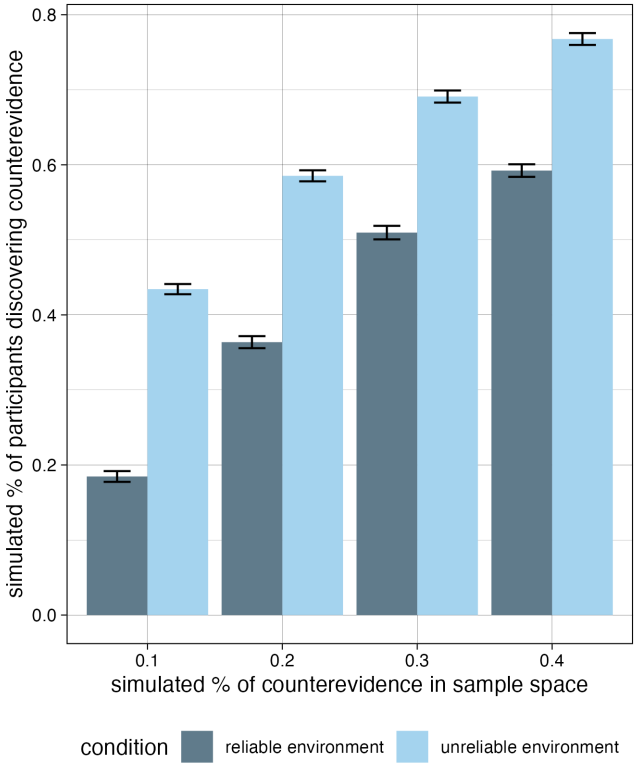

Figure 4: Simulation results reveal that children in the unreliable condition of Study 1 would have been more likely to discover counterevidence than those in the reliable condition. This pattern holds when counterevidence is both rare and relatively common. Error bars represent one SEM.

General Discussion

In order to learn accurately and efficiently, children must have an adaptive policy for deciding which claims to trust on the spot, and which to seek more evidence for. In two experiments, we investigated whether children use the reliability of their informational environment to make rational inferences about which claims warrant the most investigation. Study 1 demonstrated that children seek more evidence for a novel claim about aliens that arises in a context containing some misinformation about animals. In Study 2, children flexibly adapted their evidentiary standards according to the prior reliability of their environment. The proportion of false animal statements kids heard in the exposure phase positively predicted the amount of evidence they sampled before verifying a test claim about aliens. This was true even in a search engine environment in which each statement derived from a distinct source. Thus, children made smart inferences about their context to help them decipher how much to trust new information. They made fine-grained

assessments of the reliability of their informational environment in a known domain, inferred that this reliability would generalize to another domain, and flexibly chose a graded evidentiary standard corresponding to that reliability. Finally, we showed with a simulation that this behavior is adaptive: learners have the greatest opportunity to discover counterevidence and debunk misinformation in the most unreliable environments, where misinformation is most likely to be present.

Children's ability to calibrate their evidentiary standards to the reliability of their environments helps them confront the challenge of balancing speed and accuracy during learning. Children wasted little time verifying a claim within an environment with established reliability in a known domain. Instead, they reserved more extensive information seeking for more questionable informational environments, calibrating their evidentiary standards according to nuanced changes in reliability. While this strategy is certainly not infallible, it gives children a sensible policy for information seeking in line with resource-rational decision making (Bhui, Lai & Gershman, 2021). Even when they lack domain-relevant knowledge to judge a claim's content, they leverage sophisticated attributes of their context to guide their skepticism and exploration selectively. However, it is important to note that the scope of children's inferences–what exactly constitutes the environment that children assessed–is unclear. Future research should clarify the conceptual and temporal constraints on which experiences affect children's expectations about incoming information.

Our experiments used an open information sampling task in order to capture a graded sense of children's level of skepticism or evidentiary standards. This continuous measure allowed us to capture the quantity of evidence children searched for, which corresponded to *degrees* of belief in a given claim. This approach provides direct insight into the strength of children's belief, which is what guides their future learning and behavior. The information sampling measure we employed is also implicit, which makes it more suitable for use in younger children than explicit reports (Goupil & Kouider, 2019). This work thus builds upon literature that demonstrates that infants' and children's information seeking behavior is sensitive to uncertainty (Langenhoff, Engelmann & Srinivasan, 2023; Lapidow, Killeen & Walker, 2022; Goupil, Romand-Monnier, & Kouider, 2016). We show that information seeking is sensitive to *environmental* certainty, as well as content-specific certainty.

The context of learning matters

Children have natural skepticism We demonstrate that the knowledge that novel claims require evidence is early emerging and context-sensitive. This suggests that the most fruitful avenue of intervention may not be on skepticism itself, but on children's more specific capacities to know where to look for relevant evidence in a given domain, and to evaluate how different kinds of evidence bear on complex claims. Indeed, research suggests that children aren't sensitive to the relative strengths of explanations until early school age (Danovitch, Mills, Sands & Williams, 2021). Future research should address how, and when over the course of development, children use environmental cues to adapt their evidentiary standards in terms of the *quality* of evidence that they seek.

Early environmental experiences shape later behavior A central insight of this work is that children's approach toward novel information is shaped by expectations that are formed through experience with their informational environment. This suggests that efforts to expose children only to curated informational environments may be misguided. Early experiences with overly sanitized environments may lead children to develop overly trusting priors, which may have harmful consequences. By the same token, early exposure to more heterogeneous informational environments may allow children to "flex their skepticism muscles" and build upon their existing capacities for adaptive information seeking. More broadly, intervention efforts should focus on helping children develop a broad skill set for evaluating information, rather than attempting to control their information diets.

Acknowledgements

We thank CiCi Jiang and Jolie Witkowski for their assistance with data collection, and Steve Piantadosi and members of the Kidd Lab for helpful discussions. This work was supported by the Walton Family Foundation, Jacobs Foundation, John Templeton Foundation, and Berkeley Center for New Media. Last but not least, we are grateful to the children and families who made this research possible.

References

Baer, C., & Kidd, C. (2022). Learning with certainty in childhood. *Trends in Cognitive Sciences,* 26(10), 887–896.

Baranes, A. F., Oudeyer, P. Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in Neuroscience*, *8*, 317.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences,* 41, 15-21.

Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1088.

Common Sense Media (2019). *New Survey Reveals Teens Get Their News from Social Media and YouTube.* Common Sense Media.

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science*, *18*(6), 957-971.

Danovitch, J. H., Mills, C. M., Sands, K. R., & Williams, A. J. (2021). Mind the gap: How incomplete explanations influence children's interest and learning behaviors. *Cognitive Psychology,* 130, 101421.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761-778.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*(5), 993.

Goupil, L., & Kouider, S. (2019). Developing a reflective mind: From core metacognition to explicit self-reflection. *Current Directions in Psychological Science,* 28(4), 403-408.

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, *231*, 105325.

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences,* 113(13), 3492-3496.

Langenhoff, A. F., Engelmann, J. M., & Srinivasan, M. (2023). Children's developing ability to adjust their beliefs reasonably in light of disagreement. *Child Development,* 94(1), 44-59.

Lapidow, E., Killeen, I., & Walker, C. M. (2022). Learning to recognize uncertainty vs. recognizing uncertainty to learn: Confidence judgments and exploration decisions in preschoolers. *Developmental Science,* 25(2), e13178.

Maheshwari, S. (2017, November 4). *On YouTube Kids, Startling Videos Slip Past Filters.* The New York Times. https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html/

Mott Poll (2021). *Sharing Too Soon? Children and Social Media Apps.* C.S. Mott Children's Hospital. https://mottpoll.org/reports/sharing-too-soon-children-and-social-media-apps/

Rodriguez, A. (2018, April 26). *YouTube Kids is giving parents more control over what their kids watch.* Quartz. https://qz.com/1262977/youtube-kids-is-launching-a-mode-curated-by-humans-not-just-algorithms/