

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Learning and Believing in a Digital World

Permalink

<https://escholarship.org/uc/item/8bt0m9dm>

ISBN

9798189271465

Author

Orticio, Evan Antonio

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Learning and Believing in a Digital World

by

Evan Orticio

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Celeste Kidd, Chair

Professor Arianne Eason

Professor Jan Engelmann

Professor William Thompson

Spring 2026

Learning and Believing in a Digital World

Copyright 2026
by
Evan Orticio

Abstract

Learning and Believing in a Digital World

by

Evan Orticio

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Celeste Kidd, Chair

Our beliefs are increasingly shaped by information we encounter online, yet modern information environments like social media pose distinct epistemic challenges. How do our mechanisms of belief formation and learning meet the demands of digital environments? I present a research program identifying three key capacities that guide belief formation and learning in the digital world. First, I demonstrate that even young children adjust their fact-checking behavior according to the prior quality of their digital learning environment. Second, I show that people use the popularity of a belief as a cue to its reliability, and discuss implications for how engagement metrics on social media may impact beliefs. Third, I discuss eye-tracking findings which suggest that young adults are sensitive to the level of stimulation that is most conducive to their learning, which may help them navigate information overload online. This research begins to map a toolkit for digital epistemic vigilance, and informs strategies for fostering healthy information-seeking and critical thinking across the lifespan.

Acknowledgments

This dissertation is a product of many minds and many acts of generosity.

To Celeste: I am so grateful for your unwavering support and guidance. Thank you for creating space to be uncertain, to make mistakes, to advocate, to discover, to play. It was all a gift.

To my other committee members and mentors, Arianne Eason, Jan Engelmann, Bill Thompson, and Steve Piantadosi: Thank you for pushing me to ask deeper questions, give more cohesive talks, and make prettier plots.

To my first mentors in psychology, Stella Christie, Dan Grodner, and Molly Flaherty: Thank you for fostering my curiosity and learning during my early stages of development as a scientist. Stella, thank you for being the first person to convince me that I could really do this.

To my labmates past and present, and especially Sarah, Alex, Holly, and Carolyn: Thank you for enriching my work and my life. I'm grateful that our Berkeley Way West focus room was home to many shared hours of both focus and healthy distraction.

To my research assistants: It was a privilege to work and learn alongside you. Thank you for sharing in the joy of discovery, for asking sharp questions, and for approaching unsuspecting families at public parks to get them involved in our research.

To my friends: Victoria and Emma, thank you for making our first Berkeley apartment feel like home and for sharing this journey with me. My, thank you for being my rock. And to my other friends in Berkeley and beyond, including Bella, Eunice, Jessica, Andrew, Mina, and Sadie, thank you for being you.

To my family: Thank you for everything.

Contents

Contents	ii
1 Introduction	1
1.1 Social learning allows us to make the most of limited data	1
1.2 Learning from others requires epistemic vigilance	2
1.3 Modern information ecosystems pose distinct epistemic challenges	3
1.4 Dissertation overview	6
2 Young children adapt their fact-checking to their digital environment	8
2.1 Abstract	8
2.2 Introduction	8
2.3 Study 1	11
2.4 Study 2	15
2.5 Study 3: Simulation	18
2.6 General Discussion	21
3 Adults integrate aggregated social information in belief updating	24
3.1 Abstract	24
3.2 Introduction	24
3.3 Methods	26
3.4 Results	29
3.5 General Discussion	34
4 Young adults are sensitive to how they learn under information overload	37
4.1 Abstract	37
4.2 Introduction	37
4.3 Methods	39
4.4 Results	42
4.5 Discussion	48
4.6 Conclusion	50
5 Conclusion	52

5.1	Learning how to learn online	52
5.2	Epistemic environmentalism and the importance of trust	53
5.3	Summary	54
Bibliography		55

Chapter 1

Introduction

Humans are exceptionally efficient learners, building extensive models of the world from very little data. Yet, countless examples from the modern world appear to belie this picture. People deny the reality of climate change, distrust effective vaccines, and believe that they only use 10% of their brains. Sometimes our beliefs are unresponsive to evidence, or even update in opposite directions based on the same evidence (Cook & Lewandowsky, 2016; Lord et al., 1979), which seems to violate the fundamental principles of belief updating. How do we reconcile these apparent irrationalities with humans’ remarkable capacity for learning?

1.1 Social learning allows us to make the most of limited data

A key part of this picture is the fact that people rarely form beliefs in isolation. Most real-world beliefs are impossible to form independently because the relevant data cannot be determined by direct observation, or is otherwise indirect, ambiguous, or inaccessible to a given learner (Festinger, 1954). Thus, belief formation requires the learner to sample information from the social world. The ability to learn from others is vital. It orients us to the most relevant information, unlocks a world of knowledge that exists beyond the boundaries of our first-hand experiences, and allows us to benefit from, and contribute to, an accumulation of cultural knowledge (Alister & Perfors, 2026).

Across the lifespan, we use an intuitive understanding of how other people think and act to shape our interpretations of the evidence they generate (Gweon, 2021). Early examples of this ability involve inferring agent’s epistemic states from their actions via a naïve utility calculus (Baker et al., 2017; Jara-Ettinger et al., 2016). For example, as early as 10 months of age, infants infer the value of goals from the costs of achieving them, expecting agents to prefer rewards for which they were willing to jump farther or climb higher to reach (Liu et al., 2017). At four years of age, children infer an agent’s knowledge from their willingness to incur costs to gain information (Aboody et al., 2021). By age 7, children make even more nuanced inferences about other people’s certainty, reasoning that an agent who searches

extensively for information is less sure than one who responds immediately (Huang et al., 2019). Even when the primary evidence itself is ambiguous or inaccessible, older children can infer its meaning by observing how it affects someone else’s beliefs (Orticio et al., 2025).

Pedagogical cues facilitate even more efficient use of social information. In contexts of explicit instruction, young children expect information to be complete and highly generalizable, and use these sampling assumptions to guide richer inferences and more directed exploration (Butler & Markman, 2012; Csibra & Gergely, 2009; F. Xu & Tenenbaum, 2007). When a teacher demonstrates only one function of a toy, children infer that the toy lacks other undemonstrated functions and explore it less (Bonawitz et al., 2011). Thus, children integrate the pragmatics of communication during learning, treating a teacher’s choices of what to share as evidence about what there is to learn. This strategy is fallible but efficient: it allows learners to direct their limited attention to the most relevant information in their environment.

Our reasoning abilities extend to social interactions between multiple agents, engaging a “Theory of Minds” (Chuey & Gweon, 2025). Children temper their beliefs and explore more in the context of disagreement (Langenhoff et al., 2023, 2024), perhaps in recognition of the fact that disagreement can signal that the question is genuinely difficult or the evidence is ambiguous (Amemiya et al., 2024). Children also use consensus as a cue to the reliability of a claim (Fusaro & Harris, 2008). Impressively, they show an emerging sensitivity to how a consensus is formed. They favor independent consensus over a false consensus deriving from a single primary source (Aboody et al., 2022), and make use of affective cues to distinguish between genuine and coerced consensus (Richardson & Keil, 2022). These findings suggest that children infer key properties of evidence by reasoning about the distribution and dynamics of beliefs between different people.

1.2 Learning from others requires epistemic vigilance

Learning from socially mediated information is powerful, but comes with risks. To benefit from social information, learners must safeguard against deception or misinformation. Without vigilance, there can be no rational trust. Epistemic vigilance is underpinned by a suite of sociocognitive mechanisms to assess the reliability of the source and content of information (Sperber et al., 2010). These mechanisms include scrutiny of the source’s competence (Jarvstad & Hahn, 2011) and benevolence (Handley-Miner et al., 2023; Oktar et al., 2024), and the content’s logical consistency and coherence with prior knowledge (Hahn & Oaksford, 2007; Mercier, 2020).

Epistemic vigilance is evident from very early in development, underscoring its fundamental role in human cognition. Children attend to many relevant cues to decide whether to trust a source, including their expertise, epistemic access, and confidence (for a review, see Harris et al., 2018). When given the choice to learn from two informants about an unfamiliar topic, 4-year-olds choose the one who provided more reliable information, for example, by accurately labelling familiar words (Koenig et al., 2004). Children are also sensitive to

benevolence when determining trust: they disprefer learning from informants who are mean (Mascaro & Sperber, 2009) or trick others (Vanderbilt et al., 2011). Additionally, older children become more sensitive to others' motivations and vested interests when evaluating their claims (Heyman et al., 2007; Mills & Elashi, 2014).

Children's skepticism also extends to the content of a claim itself. By age four, children show an emerging understanding that claims should be verified, even if their sources appear reliable (Butler et al., 2020). They are skeptical of claims that are counterintuitive or inconsistent with their prior beliefs, and often devise informative ways of testing them (Cottrell et al., 2023; Hermansen et al., 2024). As they get older, they also show sensitivity to the structure of an argument, preferring non-circular and mechanistic explanations (Baum et al., 2008; Mills et al., 2019). In sum, learners integrate many sophisticated cues about the quality of testimony and the reliability of its source to trust selectively. Crucially, though, many of these cues depend on shared history and knowledge of one's informant, or significant familiarity with the topic being assessed. In more complex information environments, access to these cues is far from guaranteed.

1.3 Modern information ecosystems pose distinct epistemic challenges

While the mechanisms of epistemic vigilance established in the empirical literature are powerful within small networks, it is less clear how they scale to match the complexities of the social information that we depend on today (Mercier & Miton, 2019; Szegőfi & Heintz, 2022). In humans' evolutionary past, social learning canonically occurred within face-to-face interactions between familiar individuals. Mechanisms of epistemic vigilance are attuned to this context, in which a learner has access to relatively rich information about an informant's competence, epistemic access, and motivations. Trust is maintained because violators can incur meaningful reputational costs. By contrast, in today's large-scale societies, our success increasingly depends on information produced by large epistemic institutions, making reliability cues more opaque. Consider the trajectory of scientific data, which might be filtered through interpretation by scientists, simplification by journalists, and framing by politicians. Along this trajectory, this raw data is transformed into increasingly abstract summaries or conclusions, obscuring the underlying evidence and making verification more difficult—a process called uncertainty absorption (March & Simon, 1958; Sommer, 2026). Moreover, the collective production and dissemination of knowledge engenders an epistemic diffusion of responsibility, making it more difficult to understand whom to blame or how to enforce accountability when the information is wrong.

Relatedly, evidence for the kinds of beliefs that are important to navigating modern society is just harder to access and evaluate. Beliefs differ in the proportion of relevant evidence that is local, or directly accessible, to the learner (Sommer, 2026). Toward one end of this spectrum are perceptual beliefs, like the belief that there is a pen on my desk right

now. Virtually all the evidence I need to support this belief is accessible to me if I just open my eyes. When I go home for the day, evidence for this belief is slightly more non-local, in that it depends on my memory and the reasonable assumption from past experience that no one would move it after hours. On the opposite end of the spectrum of locality are, frankly, most of the relevant real-world beliefs. The evidence for, e.g., climate change, evolution, or the efficacy of vaccines is too vast, too dispersed, and too inferential to be accessible to an individual learner. Instead, we implicitly defer to a chain of authorities to hold these beliefs. This can explain why even people who are certain in these beliefs often cannot trace back their evidence or reasoning (Rozenblit & Keil, 2002). Evidence is highly non-local for most beliefs in domains like politics and science. That we are dealing in these domains at all is a major accomplishment of our species, and a relatively recent one. At the same time, the complexity of these domains makes us particularly vulnerable to misinformed beliefs.

Additional distortions within digital environments

The fundamental cues we use to judge information, both its source and its content, are often degraded in online environments. In face-to-face interactions, we draw on a rich combination of cues including tone, body language, and knowledge of an informant’s reputation and history of reliability. In online contexts, most of these cues are absent or easily falsified. The information we encounter on social media is often provided by anonymous or unfamiliar users whose historical accuracy, expertise, motivations, and benevolence are largely inaccessible. People can obscure or misrepresent their identities or credentials, and rarely face tangible consequences for sharing misleading or inaccurate content. The asynchronous and asymmetric nature of information exchange online compounds this problem by preventing productive back-and-forth conversations.

Moreover, the content itself can be more difficult to judge because it is often short and decontextualized, without access to the broader discourse or full space of evidence being considered. We encounter information passively from an algorithm instead of discovering it through a deliberate search. Thus, we are forced to evaluate new information without exposure to alternate perspectives or counterevidence, which we might naturally encounter in a more active inquiry. The result is a fragmented and high-entropy information environment with high barriers to careful verification.

A further distortion operates at the level of content selection. In digital environments, the volume and speed of information competing for users’ attention is unprecedentedly high (Lorenz-Spreen et al., 2019). This creates competition within what is often called an “attention economy”: social media algorithms are optimized to reward the content that maximizes engagement. These algorithmic incentives shape the kinds of content that rise to the top and, in turn, the kinds of content that are produced in the first place. Importantly, the most attention-grabbing content is not necessarily the most accurate or nuanced. Rather, factors like emotionality (Berger, 2011; Schöne et al., 2021), morality (Brady et al., 2020; Gantman & Van Bavel, 2014), and out-group animosity (Rathje et al., 2021) tend to drive engagement both online and offline. Selective attention to these dimensions is functionally adaptive in

many contexts (Cosmides & Tooby, 1992; Skowronski & Carlston, 1989), but can become a liability in social media environments designed to exploit it.

These algorithmic distortions impact our beliefs. Online misinformation strategically exploits features like emotionality and moral outrage (McLoughlin et al., 2024). Even superficial exposure to a topic online can bias how we evaluate further information. For example, merely repeating a claim makes it more believable to both adults and children (Fazio & Sherry, 2020). This illusory truth effect persists even for completely implausible claims or known falsehoods, and affects related behaviors like sharing intentions which reinforce these dynamics online (Udry & Barber, 2024). Brief, non-informational exposures to a topic in a digital setting can also bias factual beliefs in attitude-consistent ways (Orticio & Kidd, 2024).

The structure of modern information ecosystems also warps cues we use to assess social consensus. The human ability to reason about consensus is already fragile; adults often struggle to assess whether agreement reflects independent convergence or a single shared influence (Alister et al., 2025; Yousif et al., 2019). Algorithms exacerbate these issues by systematically favoring some forms of content over others, thus perverting engagement metrics as proxies for popularity or consensus. Large-scale experiments on major social media platforms like X (formerly Twitter) have demonstrated that recommendation algorithms amplify polarizing and politically extreme content (e.g., Huszár et al., 2022), meaning that the most controversial positions may be misperceived as widely held. This issue can snowball when users drift or self-select into filter bubbles or echo chambers, whose information landscapes are even less representative of the broader information ecosystem (Kim et al., 2021).

Ecological rationality

As we have seen, belief formation depends on the structure of the environment in which it occurs. This observation is central to the ecological rationality framework (Gigerenzer & Brighton, 2009; Todd & Gigerenzer, 2012), which posits that cognition cannot be characterized without reference to the environmental affordances within which it operates. Simon (1990) described rationality as the match between the computations of the mind and the structure of the environment, fitting together like two blades of a pair of scissors. A clean cut requires well matched blades. The mechanisms of epistemic vigilance that serve us nicely in small-scale interactions were calibrated to a very different environment than the one we inhabit today (Mercier & Sperber, 2017). In other words, the modern information ecosystem is a dramatically updated blade. Thus, I test how our mechanisms of learning and belief formation interact with the structural features of digital environments. This work reveals how and why seemingly adaptive learning mechanisms can be led astray online. At the same time, in the spirit of ecological active learning (Ruggeri, 2022), it showcases how learners can develop sensitivity to the structure of these environments to navigate them more effectively.

1.4 Dissertation overview

Here, I explore how our mechanisms of learning and belief formation scale up to meet the demands of modern information ecosystems, like social media and AI. Each project in this dissertation (1) identifies a key feature of modern information ecosystems that poses an epistemic challenge, and (2) highlights a capacity that learners use to address this challenge. In each case, understanding learners' limitations points toward fruitful inventions. Together, this research begins to map a toolkit for digital epistemic vigilance and informs strategies for fostering healthy information-seeking and critical thinking across the lifespan.

Chapter 2 investigates how young learners assess claims in digital environments, where sources are anonymous, varied, and decontextualized. Across two experiments and a simulation, we find that even young children scale up their fact-checking efforts according to the amount of clear misinformation they have previously encountered in a digital learning environment. Thus, children can generalize across speakers to assess the reliability of a platform (like a social media feed), and use that assessment to implement a reasonable fact-checking policy that enables the efficient detection of misinformation. These findings build upon the empirical selective trust literature in two key ways. First, they broaden the scope of what is considered a source, beyond representations of single informants. Second, our studies employ a novel, active fact-checking task which captures degrees of belief or skepticism in a single claim, rather than relying on coarser, forced-choice paradigms. This measure more accurately reflects the kinds of epistemic practices we hope to assess and foster: children directly tested novel claims, rather than accepting or rejecting them outright.

Chapter 3 asks how learners contend with new forms of social information that have become more available in modern information ecosystems. In particular, we ask how adults interpret and integrate aggregated social information, like online product reviews or engagement metrics on social media posts, which bring together the beliefs of many depersonalized voices into a single summary statistic. Across two online experiments, we provide causal evidence that when a belief is perceived to be more common, it is believed more strongly—even without presenting direct evidence for the belief. Importantly, we find that adults' beliefs changed in accordance with two predictions of a Bayesian cue integration framework, suggesting that the popularity of a belief is rationally integrated with prior knowledge. Despite its utility, I discuss how this rational belief formation mechanism can be led astray in digital environments, where algorithms can dramatically distort the perceived popularity of beliefs.

Chapter 4 examines how young adults navigate learning under information overload. In digital environments, competition within the attention economy shapes the kinds of content that is produced, with implications for learning outcomes. We test undergraduates' ability to learn under high attentional demands using "brain rot-style" (highly stimulating) content in which physics videos are presented alongside task-irrelevant, visually engaging gameplay. We find notable heterogeneity in the effect of brain rot content on learning, such that half of our sample surprisingly learned more and showed better metacognitive sensitivity when watching brain rot compared to a simpler control. Critically, learners' content preferences accurately predicted their performance across conditions. Thus, young adults are sensitive to

the level of stimulation that is most conducive to their learning, setting the stage for adaptive decision-making within environments with high attentional demands. Moreover, preference for brain rot in our task is associated with higher media consumption and tolerance for high stimulation. We discuss additional attentional signatures which point to individual variation in optimal cognitive load and provide preliminary support for an adaptive account of attentional bandwidth.

Chapter 2

Young children adapt their fact-checking to their digital environment

(Previously published: Orticio, E., Meyer, M., & Kidd, C. (2024). Exposure to detectable inaccuracies makes children more diligent fact-checkers of novel claims. *Nature Human Behaviour*, 8(12), 2322–2329)

2.1 Abstract

How do children decide when to believe a claim? Here we show that children fact-check claims more and are better able to catch misinformation when they have been exposed to detectable inaccuracies. In two experiments ($N = 122$), 4–7-year-old children exposed to falsity (as opposed to all true information) sampled more evidence before verifying a test claim in a novel domain. Children’s evidentiary standards were graded: fact-checking increased with higher proportions of false statements heard during exposure. A simulation suggests that children’s behavior is adaptive, because increased fact-checking in more dubious environments supports the discovery of potential misinformation. Importantly, children were least diligent at fact-checking a new claim when all prior information was true, suggesting that sanitizing children’s informational environments may inadvertently dampen their natural skepticism. Instead, these findings support the counterintuitive possibility that exposing children to some nonsense may scaffold vigilance towards more subtle misinformation in the future.

2.2 Introduction

Children have unprecedented access to information on their phones and computers. This fact represents a very recent shift—one of both promise and problem. The internet leaves users

exposed to unprecedented amounts of misinformation. Exposure to misinformation can lead to the long-term adoption of false beliefs in both adults (Brown & Nix, 1996) and children (Fazio & Sherry, 2020). This is true even when the learner is aware of this bias (Fazio et al., 2015). Misinformation exposure is also expected to increase with the widespread adoption of generative AI models like ChatGPT and Bard. When these models produce fabricated information in their outputs, they can transmit them to users (Kidd & Birhane, 2023). Children are likely most vulnerable because they have less world knowledge (S. Xu et al., 2022) and are biased to trust information (Jaswal et al., 2010), particularly under conditions of uncertainty (Plate et al., 2021). Indeed, even when preschoolers directly observe data that conflicts with testimony, they rarely seek additional data and struggle to disregard the misleading information (Hermansen, Ronfard, Harris, & Zambrana, 2021). Despite these unique vulnerabilities, the overwhelming majority of work on misinformation centers adults (e.g., Ecker et al., 2022).

What we know of children’s media habits suggest that they are immersed. A third of American children are on at least one social media platform by age 9 (C.S. Mott Children’s Hospital, 2021). A majority of American teens get their news from social media or YouTube (Common Sense Media, 2019). And children who have used ChatGPT for schoolwork report using it in place of traditional search (Common Sense Media, 2023). Thus children’s media diets are rife with dubious sources. How do we best prepare children to navigate this complex informational sea?

The preeminent solution has been to shield children from misinformation via sanitized platforms. YouTube Kids, for example, curates a small selection of child-focused content through a combination of automated filters and human review (Rodriguez, 2022). This solution is limited by its reactive nature. As an example, YouTube Kids received widespread criticism when a Guardian article reported on a multitude of videos featuring themes that were not appropriate for children (e.g. violent and sexual situations) which were inaccurately labeled as child-friendly by the platform’s filters, likely because they contained characters from children’s movies and shows (“Elsagate”, Maheshwari, 2017; Subedar & Yates, 2017). Efforts to sanitize content for children are resource-intensive and subjective (who decides what is age-appropriate?). Moreover, automated curation approaches are easily gamed, and human curation approaches cannot scale as rapidly as new content is produced (Kallioniemi, 2021).

Another proposed strategy for safeguarding people from misinformation comes from inoculation theory (for reviews, see Compton et al., 2021; Lewandowsky & Van Der Linden, 2021; Van Der Linden, 2022). Inspired by an analogy to biomedical inoculation, the theory postulates that preemptively exposing learners to a weakened form of a misleading argument can confer immunity to its persuasiveness later on. This process involves ‘prebunking’ the argument by refuting false information in advance, and/or deconstructing misleading argumentation techniques more broadly. Researchers claim to have successfully inoculated adults against misinformation spanning many topics, including climate change, vaccination, and extremist ideology (Basol et al., 2020; Braddock, 2022; Van der Linden et al., 2017). However, inoculation interventions are fragile, ephemeral, and difficult to scale (Maertens et

al., 2021). Inoculation interventions are short-acting, and ineffective after 48 hours without direct and immediate reinforcement (Capewell et al., 2024). These interventions have also been criticized for methodological weaknesses in the assessment of their efficacy (Chan & Albarracin, 2023; Guay et al., 2023; Williams, 2023). For example, a recent analysis found no evidence that inoculation improves discrimination, but rather that it induces a potentially counterproductive, negative response bias (Modirrousta-Galian & Higham, 2023).

Further, misinformation inoculation techniques are rarely tested in young children—and there’s reason to expect they may not achieve even the modest, ephemeral effects seen in adults because of differences in children’s decision-making and metacognition. Children have less developed metacognitive skills than adults, a tendency toward overconfidence (Finn & Metcalfe, 2014; Lipko et al., 2009; Salles et al., 2016), and less executive function (Best & Miller, 2010). Thus, it may be more difficult to find interventions that effectively lower overconfidence and slow decision-making about factual accuracy for children.

Despite this, research suggests that children can accurately assess the epistemic quality of human sources (see Harris et al., 2018 for a review, and Tong et al., 2020 for a recent meta-analysis). When given the choice to learn from two informants about an unfamiliar topic, 4-year-olds choose the one who provided more reliable information, e.g., by accurately labeling familiar words (Koenig et al., 2004). Similarly, children this age will endorse a statement from a knowledgeable informant over a conflicting statement from one who is less so (e.g., Corriveau & Kurkul, 2014; Danovitch & Alzahabi, 2013). These assessments of informants are nuanced. Children as old as four endorse statements from informants in accordance with the proportion of inaccurate statements they make (e.g., 80% versus 20%), demonstrating that judgements of trust extend beyond a binary of trustworthy or untrustworthy (Pasquini et al., 2007). Furthermore, children change their endorsement of an informant’s statement if later empirical checks show it to be inaccurate (Hermansen, Ronfard, Harris, Pons, & Zambrana, 2021).

Children may leverage this capacity to differentiate between high- and low-quality sources in digital media environments. However, two issues still limit our understanding of how children might fare in these more complex informational ecosystems. For one, the sources on the internet are seldom transparent and often cannot be traced back to a particular agent. Rather, the statements children encounter derive from something more nebulous: webpages, posts, or a series of such entities supplied by an algorithm. It is unclear how children will perform in such cases when informant-specific cues are unavailable. Secondly, measurements of selective social learning have traditionally relied on a forced-choice between rejecting or endorsing statements from a given source. In reality, trust is more nuanced and has downstream consequences on what information people choose to seek out or not. Given uncertainty regarding a novel statement, children may opt to fact-check it. Fine-grained evaluation of children’s investigatory response would provide a more informative picture of their skepticism. Moreover, investigation rather than immediate rejection is a solid strategy for the effective use of a source that provides partial misinformation.

Here, we propose an approach for motivating adaptive skepticism toward digital misinformation which builds upon children’s known capacities to use the statistical properties

of information in their environment—capacities for which we have strong evidence even in infancy (Fiser & Aslin, 2002; Saffran et al., 1996; F. Xu & Garcia, 2008). The idea leverages the fact that children attend to statistical properties of their environments in order to form expectations which then modulate their learning and behavior (Kidd et al., 2012; Saffran et al., 1996). Existing empirical work shows that children wait longer in a delay-of-gratification task when given evidence that waiting will pay off (Kidd et al., 2013; Watts et al., 2018). We hypothesize that, in a similar fashion, children will use the prior reliability of information in a given context to adjust their a priori skepticism toward new claims. In two experiments and a simulation, we test whether controlled but imperfect informational environments may serve as useful scaffolding for children’s abilities to detect misinformation. Exposure to detectable inaccuracies may provide critical opportunities for children to express their skepticism and practice key critical thinking skills.

2.3 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.

Methods

Participants

Sixty 4- to 6-year-old children ($M_{age} = 5.51, SD = 0.89$, 47% White, 58% female, 40% male, 2% gender not declared) were recruited from parks in the San Francisco Bay Area. Three additional children were excluded from analysis because they had watched another child participate or were too distracted to complete the study. Children received a small toy valued at \$1-2 as compensation for their participation.

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).

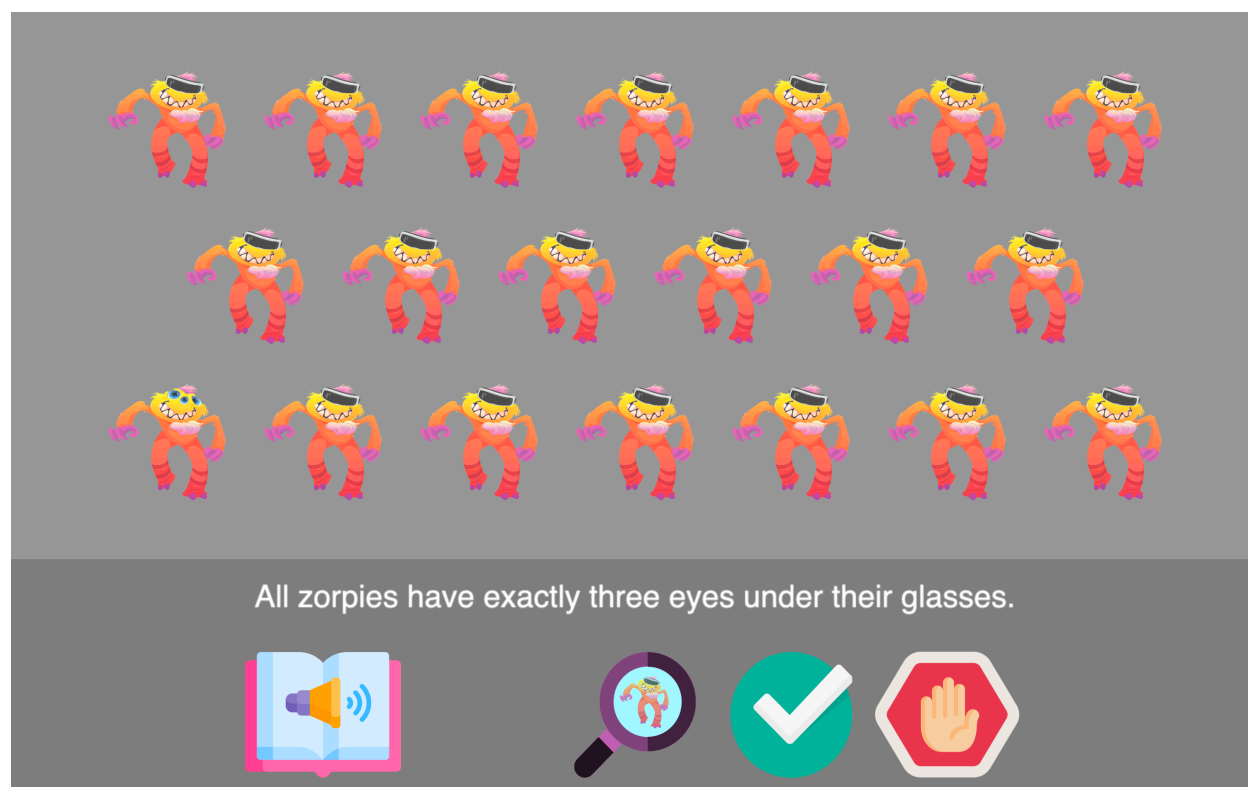


Figure 2.1: Test phase, identical in Studies 1 and 2. After checking a zorpie (e.g., bottom row, first from left), children could choose to accept the statement (green button), reject the statement (red button), or check another zorpie before deciding (magnifying glass). Illustrations and icons were adapted from freepik.com and flaticon.com.

The facts varied by condition (between-subjects, $n=30$ per condition). 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. The first two trials were considered practice trials, and children were given feedback if they were wrong. No feedback was provided on the remaining 10 trials.

In the subsequent test phase (see Figure 2.1), 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, $t(29) = 21.88$, two-sided $p < .001$, Cohen's $d = 3.99$, $95\%CI = (8.99, 9.81)$) and unreliable conditions ($M = 8.43$ of 10 correct, $t(29) = 8.64$, two-sided $p < .001$, $d = 1.58$, $95\%CI = (7.62, 9.25)$), indicating that we successfully manipulated the perceived reliability of information in the exposure phase. Nine of the 60 participants failed to achieve 80% accuracy, but their exclusion does not affect any results, so we retain 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.

Children seek more evidence in unreliable environments

Children increase their standards of evidence for new claims in an environment containing some misinformation. Figure 2.2 shows the number of zorpies children sampled before deciding to accept or reject the test claim by condition. We used the `rstatix` package in R to run a Wilcoxon paired signed rank test assessing the effect of condition on the amount of evidence sampled. On average, children in the Unreliable condition sampled more evidence than those in the Reliable condition ($M = 6.63$ vs. 2.17 zorpies, location parameter $= -1$, two-sided Wilcoxon $W = 233$, $p < .001$, $r = 0.43$, $95\%CI = (-2, -1)$). When exposed to some misinformation, children sought out significantly more evidence 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 information 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

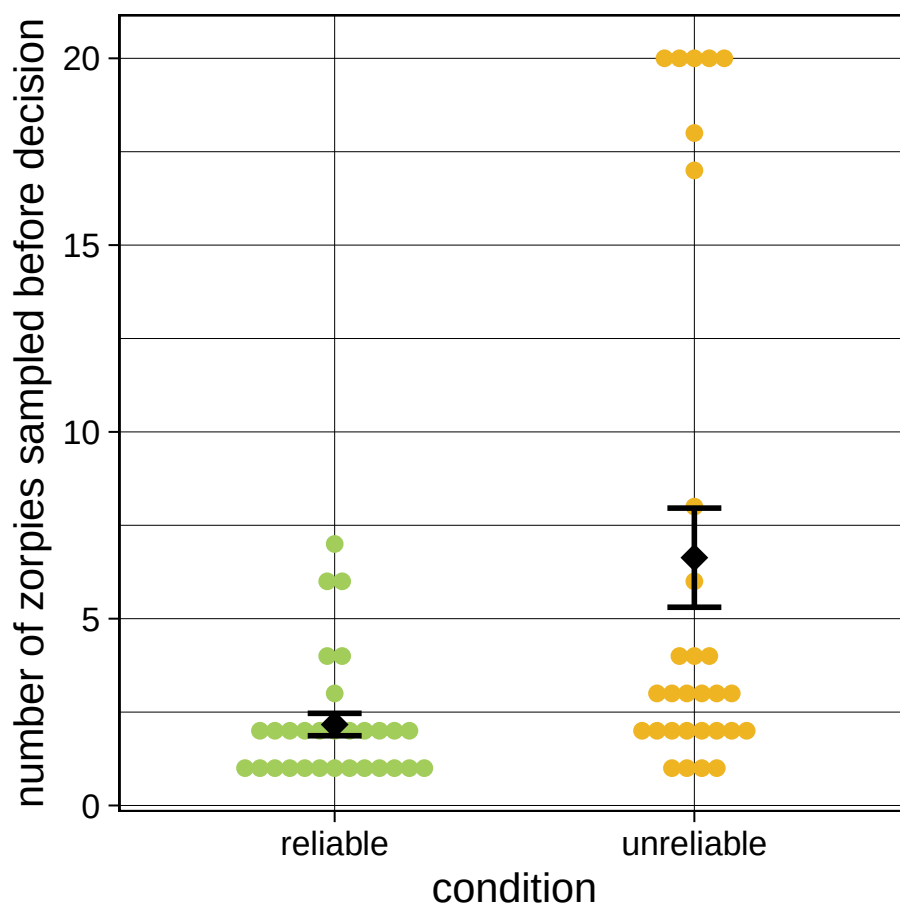


Figure 2.2: Children sampled more evidence in the unreliable condition ($n = 30$ per condition). Dots are individual data points, diamonds are condition means, and error bars represent one SEM. The effect of condition remains robust after winsorization, ensuring that the highest sample values do not drive the effect.

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 (location parameter = -1 , two-sided Wilcoxon $W = 233$, $p < .001$, $r = 0.43$, $95\%CI = (-2, -1)$), 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 betabin function from the aod package in R to run a beta-binomial regression using condition and standardized age to predict the proportion of zorpies sampled. This analysis replicated the main effect of the unreliable condition ($\beta = 0.84$, $z(55) = 2.78$, $p = .005$, $95\%CI = (0.25, 1.43)$), but revealed no main effect of age ($\beta = 0.07$, $z(55) = 0.30$, $p = .764$, $95\%CI = (-0.37, 0.50)$) and no interaction ($\beta = -0.02$, $z(55) = -0.07$, $p = .942$, $95\%CI = (-0.61, 0.56)$). There was no evidence of changes in how children responded to the reliability of information from ages 4 through 6 in our sample.

2.4 Study 2

The reliability of a body of information is not all-or-nothing. Do children appreciate nuances in the reliability of their broader informational 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 assumed that the information in Study 1 came from a single, cohesive source. The task was framed as an ebook, and children heard all statements in audio recordings using the same voice. Can children still make smart inferences about claims derived from 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 informational environment to adapt their information seeking.

Methods

Participants

Sixty-two 4- to 7-year-old children (Mage = 5.88, SD = 1.06, 52% White, 40% female, 50% male, 10% other gender or undeclared) were recruited from parks in the San Francisco Bay Area. Participants were pseudo-randomly assigned to five between-subjects conditions with sample sizes of $n = 12, 13, 13, 12$, and 12 , respectively. 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. Children received a small toy valued at \$1-2 as compensation for their participation. Data collection stopped at the end of a testing day when the sample size had reached or surpassed 60 children, which was the sample size of Study 1.

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. Exposure trials were presented in a randomized order. 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, $t(61) = 33.3$, two-sided $p < .001$, $d = 4.23$, $95\%CI = (8.99, 9.50)$). Four of the 62 participants failed to achieve 80% accuracy, but their exclusion does not affect any results, so we retain 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).

Graded sensitivity to reliability

Figure 2.3 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 beta-binomial regression revealed that the proportion of false statements from the exposure phase positively predicted the proportion of the 20 zorpies that participants sampled in the test phase ($\beta = 1.10$, $z(59) = 2.04$, $p = .041$, $95\%CI = (0.04, 2.16)$). Skepticism increased with increases in the number of false statements in the exposure phase, manifesting in more extensive information search in the test phase. A linear model finds the same effect and supports the same conclusions. 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 context 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.

While we observed that increases in the proportion of misinformation led to increases in the sampling of evidence, it is unclear whether this would generalize across all tasks. In

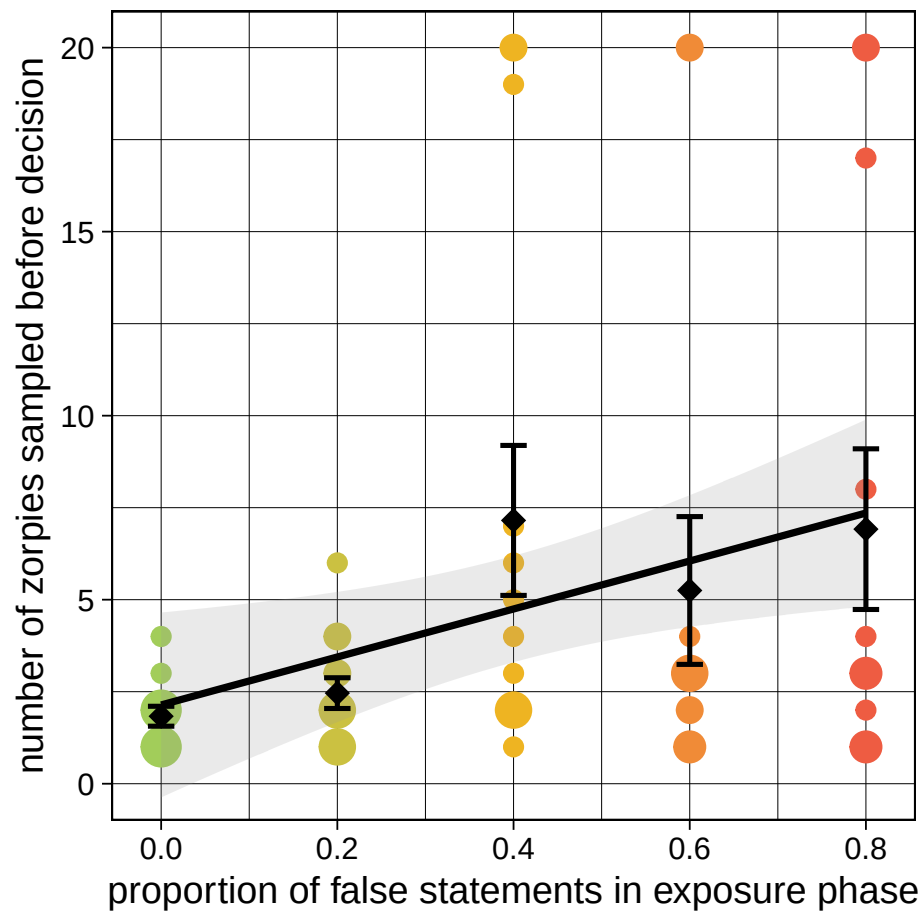


Figure 2.3: Children ($N = 62$) sampled more evidence as the reliability of their environments decreased. Amount of evidence sampled (out of a possible 20 zorpies) vs. proportion of false statements in exposure phase in Experiment 2. The size of the dot represents the number of data points. Diamonds are conditional means and error bars represent 1 SEM. Line is the linear regression fit with a 95% confidence interval.

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 found linear associations between low certainty and information seeking in children (Coughlin et al., 2015) 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 et al., 2014; Wang et al., 2021). 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 (Baer & Kidd, 2022; Goupil & Proust, 2023). 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 effect we observe may therefore be the combination of two patterns of responses: a monotonic, positive relationship dampened by a subgroup exhibiting a U-shaped pattern. Alternatively, the pattern may be influenced by variations in children’s metacognition or executive function.

Sensitivity may increase with age

Does children’s skepticism become more finely attuned to the reliability of their informational environment as they age? We ran a beta-binomial regression using standardized age and the standardized proportion of false statements in the exposure phase to predict the proportion of zorpies sampled. First, we replicated the main effect of environmental reliability (i.e., proportion of false statements in the exposure phase, $(\beta = 0.37, z(57) = 2.48, p = .013, 95\%CI = (0.08, 0.67))$). In addition, this analysis revealed a significant main effect of age $(\beta = 0.43, z(57) = 2.81, p = .005, 95\%CI = (0.13, 0.73))$ and a significant interaction $(\beta = 0.35, z(57) = 2.14, p = .032, 95\%CI = (0.03, 0.67))$. 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 they had encountered a high proportion of false information in the past.

2.5 Study 3: 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 deter-

mine whether experience learning in an unreliable environment enables children to identify misinformation more easily.

We ran four simulations of 100,000 hypothetical experiments each, in which a randomly sampled proportion of the zorpies in the sample space served as hypothetical counterevidence to the test claim (i.e., zorpies without three eyes). Then, using participants' real sampling behavior from Study 1, we computed the proportion of times each individual would have successfully found one or more of the counterevidence zorpies during the task across the 100,000 simulated experiments. This proportion represents the simulated likelihood of a given participant to discover counterevidence for the claim and, in doing so, debunk misinformation in the test phase. We repeated this simulation procedure for four proportions of counterevidence in the sample space such that 10%, 20%, 30%, and 40% of the available zorpies represented evidence against the test claim. Figure 2.4 visualizes the mean likelihood of discovering counterevidence across the 100,000 simulations, according to the participant's experimental condition from Study 1 and the simulated proportion of counterevidence within the sample space.

To test whether children in the unreliable condition were more likely to discover counterevidence during sampling, we ran a beta-binomial regression using experimental condition (reliable vs. unreliable) to predict the proportion of simulation runs in which the participants discovered one or more pieces of counterevidence during sampling. Since we used participants' actual sampling behavior from Study 1, there is dependence within each participant's outcomes at each of the four simulated proportions of counterevidence. Therefore, we analyzed the results of the four simulations (with 10%, 20%, 30%, and 40% counterevidence in the sample space) separately with identical beta-binomial models. Results indicated a significant effect of condition in the 10% counterevidence condition ($\beta = 1.40, z = 4.05, p < .001, \phi = 0.47, 95\%CI = (0.72, 2.08)$), the 20% counterevidence condition ($\beta = 1.73, z = 5.06, p < .001, \phi = 0.41, 95\%CI = (1.06, 2.40)$), the 30% counterevidence condition ($\beta = 1.72, z = 5.31, p < .001, \phi = 0.36, 95\%CI = (1.08, 2.35)$), and the 40% counterevidence condition ($\beta = 1.63, z = 5.24, p < .001, \phi = 0.32, 95\%CI = (1.02, 2.23)$), such that participants in the unreliable condition discovered counterevidence in a higher proportion of simulations compared to those in the reliable condition. Thus, across varying proportions of counterevidence in the sample space, children who were exposed to more misinformation in the exposure phase were more likely to discover counterevidence about a novel, verifiable claim in the subsequent test phase (if such counterevidence existed). The simulation results replicate the expected probabilities of each participant sampling one or more pieces of counterevidence without replacement based on the hypergeometric distribution. These convergent findings support the commonsense conclusion that unreliable informational environments elicit increased skepticism and fact-checking behavior, which enables children to debunk misinformation more readily.

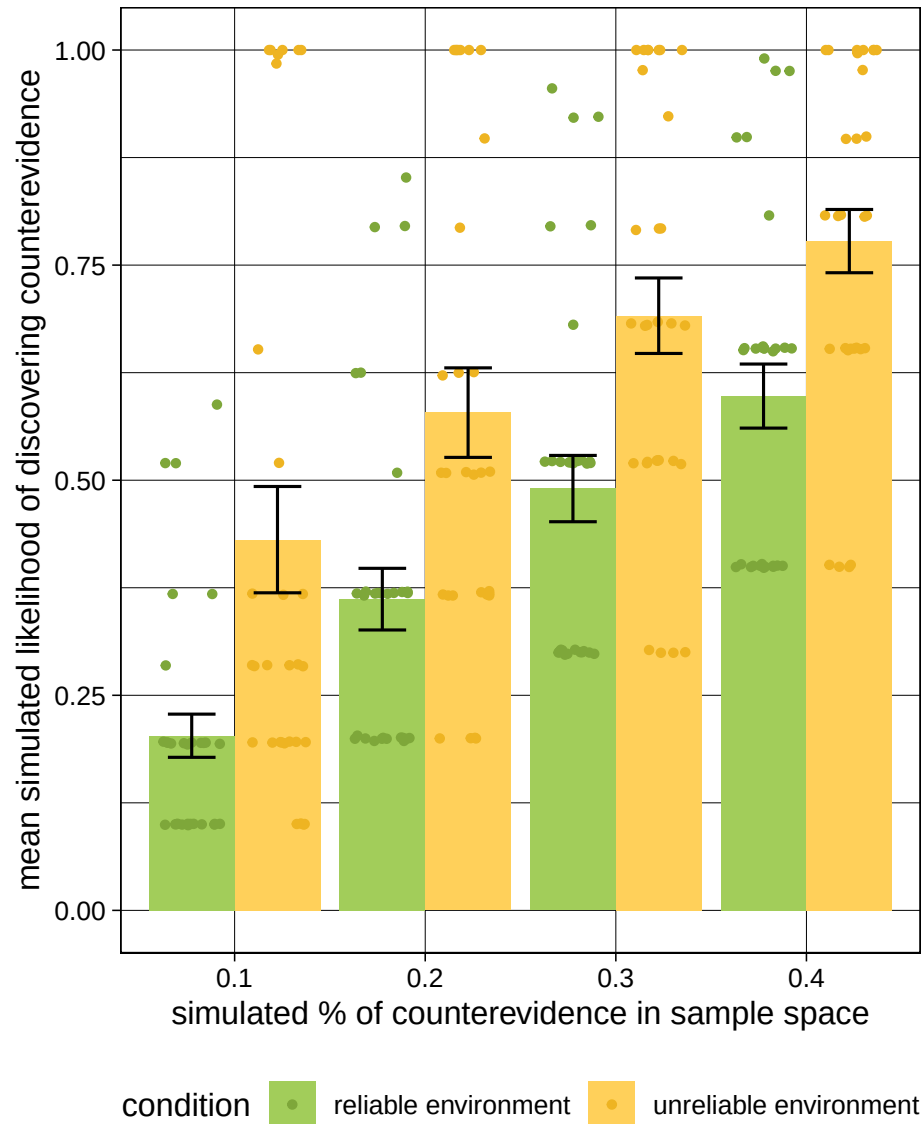


Figure 2.4: Simulation results reveal that children in the unreliable condition of Study 1 ($n = 30$ per condition for each simulation) would have been more likely to discover counterevidence than those in the reliable condition. This pattern holds when counterevidence is both rare and relatively common. Dots are likelihoods for each participant in Study 1. Error bars represent 1 SEM.

2.6 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 whether a new claim warrants fact-checking. Study 1 demonstrated that children seek more evidence for a novel claim about aliens that arises in a context containing some misinformation about animals. Study 2 showed that this evidence seeking behavior increased in proportion to the number of false claims children were previously exposed to. Moreover, this effect held when information was presented in a search engine context, in which each claim derived from a distinct source. Children thus adjusted their level of skepticism not in accordance with speaker-specific cues but with the reliability of a broad informational environment. They made fine-grained assessments of the reliability of incoming information in a known domain, inferred that this reliability would generalize to another domain, and chose a graded evidentiary standard corresponding to that reliability. Finally, we showed with a simulation (Study 3) 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 contexts, modulating 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 et al., 2021). Even when they lack domain-relevant knowledge to judge a claim’s content, children leverage sophisticated attributes of their context to guide their skepticism and exploration selectively. These findings dovetail with recent work showing that adults’ beliefs about naturalistic news headlines adapt to the prior base rate of true headlines in their news feeds (Orchinik et al., 2025), which emphasizes the relevance of this environmental adaptation for understanding real-world beliefs.

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 confers several advantages over the forced-choice paradigm that is common in the selective trust literature. The continuous measure indexes the strength or uncertainty of children’s individual beliefs, rather than demonstrating only a relative belief in one informant’s testimony over another’s. It also provides insight into children’s strategies for translating their level of trust into rational information-seeking behavior, which is crucial in order to arrive at true beliefs in complex environments. Further, the measure we employed is 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 (Goupil et al., 2016; Langenhoff et al., 2023; Lapidow et al., 2022). We show that information seeking is sensitive to environmental certainty, as well as content-specific certainty.

Let us consider some limitations. While these data demonstrate childrens’ sensitivity to their informational environments, it is unclear precisely how they conceptualize the information source in the first place. Children may have directed their skepticism toward the content of the particular ebook (Study 1) or search engine (Study 2) they were exposed to, toward any information presented on the touchscreen computer, or toward any information presented by the experimenter, among other possibilities. The present experiments cannot distinguish between these possibilities, but evidence suggests that children of this age can limit the scope of their inferences at least to a particular technological device. Preschoolers and kindergarteners express selective trust for computers (Danovitch & Alzahabi, 2013), social robots (Brink & Wellman, 2020), and internet sources (Tong et al., 2022) based on their past accuracy. Still, future work should clarify the limits of children’s skepticism in this paradigm by manipulating whether the test phase occurs in the same context as the exposure phase.

The nature of the misinformation that children encounter likely impacts the skepticism they express. In our studies, misinformation was maximally detectable: errors in the animal facts were blatant and contradicted the accompanying pictures, which were always consistent with reality. It was important to establish that pictures in the environment were trustworthy because pictures served as the evidence to be sampled in the test phase. However, these elements of our task are unrepresentative of typical misinformation that children might encounter online. Misinformation is often more subtle, and the ground truth is rarely directly accessible. If misinformation is undetectable, it likely won’t impact children’s subsequent skepticism and fact-checking behavior. What’s more, the errors in our task were related to key properties of the animals (e.g. “Zebras have red and green stripes”). Children are more vigilant against speakers who make semantic errors like these compared to speakers who make episodic errors, potentially because the former “resist folk-psychological explanation” (Stephens & Koenig, 2015). Although it’s unclear whether children make the same folk-psychological inferences about these broader, depersonalized informational environments, there is evidence that preschoolers use the accuracy of text-based sources to make inferences about their authors (Vanderbilt et al., 2018). At the same time, children tend to attribute a computer’s errors to problems with the device itself rather than with its human user (Danovitch & Alzahabi, 2013). Thus, it remains unclear whether children’s evidentiary standards would scale up in the same way in response to other forms of misinformation in digital contexts. Still, the fact that children’s skepticism shows graded adaptation to the rate of misinformation in an environment composed of many distinct voices suggests that these inferences need not be tethered to a representation of a single agent. Future work should assess how qualitatively different forms of misinformation correspond to children’s environment-level skepticism.

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 and rob them of opportunities to develop critical thinking skills. 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. This idea is consistent with evidence that exposing adults to blatant misinformation makes them less susceptible to more subtle misinformation compared to a control condition (Loftus, 1979), although recent direct and conceptual replications failed to find this effect (O’Donnell & Chan, 2025).

Crucially, the form of skepticism engendered by our paradigm was not overhasty rejection of all novel information, but critical assessment of that information. This stands in contrast to work which suggests that adaptation to a largely false news feed causes adults to disproportionately misidentify true headlines as false (Altay et al., 2025; Orchinik et al., 2025). Real world circumstances often justify this kind of adaptation. For example, a long history of medical discrimination and abuses like the Tuskegee Syphilis Study are justified reasons for adaptive skepticism that explain why Black Americans are more likely than other racial groups to endorse health-specific COVID-19 conspiracy beliefs (A. C. Smith et al., 2022) and to refuse vaccination (Majee et al., 2023). However, given that most media consumed by the public is true (Acerbi et al., 2022), an overgeneralized resistance toward reliable information has potential to harm people’s discrete beliefs and to erode trust in public institutions more broadly. In our study, when presented with misinformation, children did not show any indication that they would immediately discard true claims from a novel domain. Instead, when uncertain of a claim’s truth, children opted to perform a sometimes tedious empirical check of its veracity. In preparing children for an informational environment where misinformation is present but not omnipresent, we contend that the critical engagement with novel claims fostered by our paradigm is preferable to mere acceptance or rejection. Operating in environments where detectable misinformation is present but checkable seems to be a promising means of scaffolding this analytic disposition in children.

Overall, these findings demonstrate that the knowledge that novel claims require evidence is early emerging and context-sensitive. Even so, skepticism is only a useful signal if children know how to act upon it in rational and informative ways. While our task provides the full space of evidence to verify the test claim (in the form of the 20 zorpies), fact-checking in the real world is significantly more complex. 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 et al., 2021). In sum, intervention efforts should focus on helping children develop a broad skill set for evaluating information, rather than attempting to control their information diets.

Chapter 3

Adults integrate aggregated social information in belief updating

(Previously published: Orticio, E., Martí, L., & Kidd, C. (2022). Social prevalence is rationally integrated in belief updating. *Open Mind*, 6, 77–87)

3.1 Abstract

People rely on social information to inform their beliefs. We ask whether and to what degree the perceived prevalence of a belief influences belief adoption. We present the results of two experiments that show how increases in a person’s estimated prevalence of a belief led to increased endorsement of said belief. Belief endorsement rose when impressions of the belief’s prevalence were increased and when initial beliefs were uncertain, as predicted by a Bayesian cue integration framework. Thus, people weigh social information rationally. An implication of these results is that social engagement metrics that prompt inflated prevalence estimates in users risk increasing the believability and adoption of viral misinformation posts.

3.2 Introduction

The amount of data required to construct all of our beliefs “from scratch” is intractable given the limitations of our attention and the complexity of the world. Additionally, most truths cannot be determined by direct reference to the physical world anyway (Festinger, 1954). Instead, relevant evidence remains inaccessible to the individual for most real-world beliefs and must be mediated through other agents (Perfors & Navarro, 2019). Belief formation thus depends on information sampled from the social world. This process is distinct from normative social influence, in which conformity to a group norm takes precedence over conflicting private information about the ground truth (Deutsch & Gerard, 1955). Rather, this social sampling is necessary for acquiring information that is complex or otherwise unavailable to the learner.

Empirical evidence makes clear that social information affects beliefs, even while it leaves open how and why specific social information is integrated. This work dates back to Asch (1951), who demonstrated that people can report erroneous perceptual judgments when influenced by inaccurate information from confederates. Likewise, appeals to expert consensus effectively promote belief change, for example, for anthropogenic climate change (Goldberg et al., 2020; Van der Linden et al., 2017), nuclear power (Kobayashi, 2018), GMOs (Kerr & Wilson, 2018), and vaccination (Van Prooijen et al., 2015). These social influences could be the result of social pressures to conform to others, especially those of higher status. Van der Linden et al.’s Gateway Belief Model (2019) argues instead that social information simply licenses belief revision. Alternatively, a rational account would predict social information should be integrated with the person’s a priori beliefs in ways that are sensitive to the reliability of both information sources.

Evidence in support of rational accounts of social information integration includes that individuals find a larger number of people more convincing than a smaller number (e.g., Bond, 2005). Relatedly, the number of endorsers of a belief is known to be a key influencer of belief adoption. For example, Ransom et al. (2021) found that the raw number of people endorsing a view was a more important predictor of belief adoption than the diversity of arguments made. Some evolutionary models of social learning also have a rational flavor, suggesting intelligent organisms sometimes learn socially by adopting a copy-when-uncertain strategy. Learners copy the behavior of others when they have no relevant private information (Kendal et al., 2009; Rendell et al., 2010; Toelch et al., 2014). This evidence is consistent with the rational uptake of social information into one’s behaviors, but examines cases of discrete choice rather than graded integration of new social information with held beliefs.

Informational social influence as cue integration

The role of social information in belief formation may be better understood within a cue integration framework, borrowed from the perception literature (e.g., Ernst & Banks, 2002; Knill & Richards, 1996). Observers combine signals from different perceptual senses to arrive at a unified percept. In forming a new belief, a learner similarly integrates private information, like domain-relevant knowledge, with social information, like the prevalence of a belief in a sampled population. Importantly, each cue is flexibly weighted according to its estimated reliability. Strong reliance on social information can be rational, especially under conditions of uncertainty. Within this framework, even the conformity to the majority in Asch’s (1951) line-matching experiments could reflect optimal information integration rather than—or in addition to—normative pressure to conform. If the participant assumes that others are unbiased and have low error rates, as would reasonably be expected in simple perceptual tasks, then the probability of the participant being correct given the confederates’ convergence on a different answer is low (Toelch & Dolan, 2015). Thus, an optimal Bayesian learner would give high weight to the social information.

The role of social information in empirically unsupported belief

These principles may also explain viral cases of online misinformation. Much of the existing literature on conspiracy and pseudoscientific beliefs has sought to explain them away by appealing to ancillary influences on belief formation, like emotionality (Vlasceanu et al., 2020), political extremism (Van Prooijen et al., 2015), inattentiveness (Pennycook & Rand, 2019), and bullshit receptivity (Pennycook & Rand, 2020). However, these arguments fail to acknowledge that many online platforms have created an environment that makes rational belief formation processes highly susceptible to misinformation. False news has been found to spread “farther, faster, deeper, and more broadly” than true news online (Vosoughi et al., 2018), due in part to the formation of echo chambers. Given the high-virality potential of online misinformation, it is likely that people encounter high social engagement metrics, like the number of likes and shares, accompanying false claims. These metrics may serve as a cue to a belief’s prevalence in the population, which may in turn impact the user’s own impression of the belief’s legitimacy. This information may be particularly problematic when it is encountered while a person is “doing their own research” and thus in a state of high uncertainty.

Overview of present research

We conduct two experiments to test whether increasing the perceived social prevalence of a belief increases its believability in the absence of direct evidence. Experiment 1 tests this hypothesis using a set of current, real-world conspiratorial and pseudoscientific beliefs. We further test whether prevalence information is integrated rationally as predicted by a cue integration framework. If so, the social prevalence cue should be weighted against existing evidence such that it elicits the strongest belief change (a) when new data is most convincing, that is, estimates of the belief’s prevalence change the most in response to the data, and (b) when the initial belief is most uncertain. Experiment 2 replicates Experiment 1 but implements a cover task to reduce possible demand characteristics. In the next section, we present the methods for each experiment and then discuss the results of both together.

Understanding how social prevalence estimates impact belief adoption is important for understanding how to develop interventions against echo chambers online. The ubiquity and prominence of inflated cues to the social prevalence of beliefs (e.g., via visible metrics of likes, shares, or views) may increase engagement while at the substantial cost of leading people to be more likely to adopt beliefs they would not under less artificial circumstances.

3.3 Methods

Experiment 1

Participants read a series of statements relating to real-world conspiratorial, pseudoscientific, or misinformed beliefs (e.g., “The earth is flat,” “Wearing masks is harmful to the health

of the mask wearer,” “The U.S. government planned the 9/11 attack on the World Trade Center”). We chose beliefs that are recently attested but thinly and weakly held (Marti et al., 2021). These low-probability beliefs serve as a stringent test of the strength of social influence. On each trial, participants provided (1) a likelihood estimate of the belief (“How likely do you think it is that the statement is true?” on a 0–100 slider scale), and (2) a prevalence estimate (“How many people out of 100 do you think believe the statement is true?”). The experiment began with three simple practice trials involving statements of fact (e.g., “Plants need water to grow”) with feedback. We included four more trials of the same type in the main task without feedback as attention checks, for a total of 37 trials.

Participants repeated the same trials in a shuffled order in a second block. However, before evaluating the likelihood and prevalence of each statement, participants viewed a sample of data indicating how many of 10 survey respondents believed the shown sentence (see Figure 3.1). The sample for the 10 people approximately matched the participant’s own estimate for the prevalence of the belief in block 1 (Control condition) on half of the trials. On the other half of trials, the number of people in the sample endorsing the pseudoscientific or conspiratorial belief was 40% higher than the participant’s initial prevalence estimate (Higher Prevalence condition). For example, if a participant estimated in block 1 that 19 of 100 people believe the given statement is true, then in block 2 they would be shown that 2 of the 10 survey respondents believe the statement (Control condition) or 6 of the 10 believe it (Higher Prevalence condition). Note that, for variation, some statements were related to an attested conspiracy or pseudoscientific belief but worded in their inverse (true) forms, for example, “Humans have landed on the moon.” In this case, the prevalence of the sample was 40% lower, that is, in the direction of the empirically unsupported belief, in the Higher Prevalence condition. These items are reverse coded in our analyses, so we retain the Higher Prevalence label for simplicity. The participants then gave estimates for the prevalence and likelihood of the belief.

We debriefed participants at the end of the experiment. We provided a representative prevalence estimate of all of the beliefs based on a large sample of over 900 Americans (Marti et al., 2021) and reminded participants that prevalent beliefs are not necessarily true.

Experiment 1 guarded against the risk of participants feeling pragmatic pressure to change their prevalence and belief estimates by nature of being asked twice (e.g., demand characteristics) in two ways. First, we collected ratings for a large number of statements (30) in shuffled orders to make remembering what had been asked previously difficult. Second, we asked participants after the experiment what they believed the experiment to be about in order to determine whether participants guessed the purpose and may have altered their responses accordingly. Experiment 2 adds the additional guard of a cover task.

Experiment 2

Experiment 2 replicates Experiment 1 with one change: before block 2, participants were told that the prevalence data was part of a memory task. This cover task was intended to provide an additional guard against demand characteristics. We instructed participants to

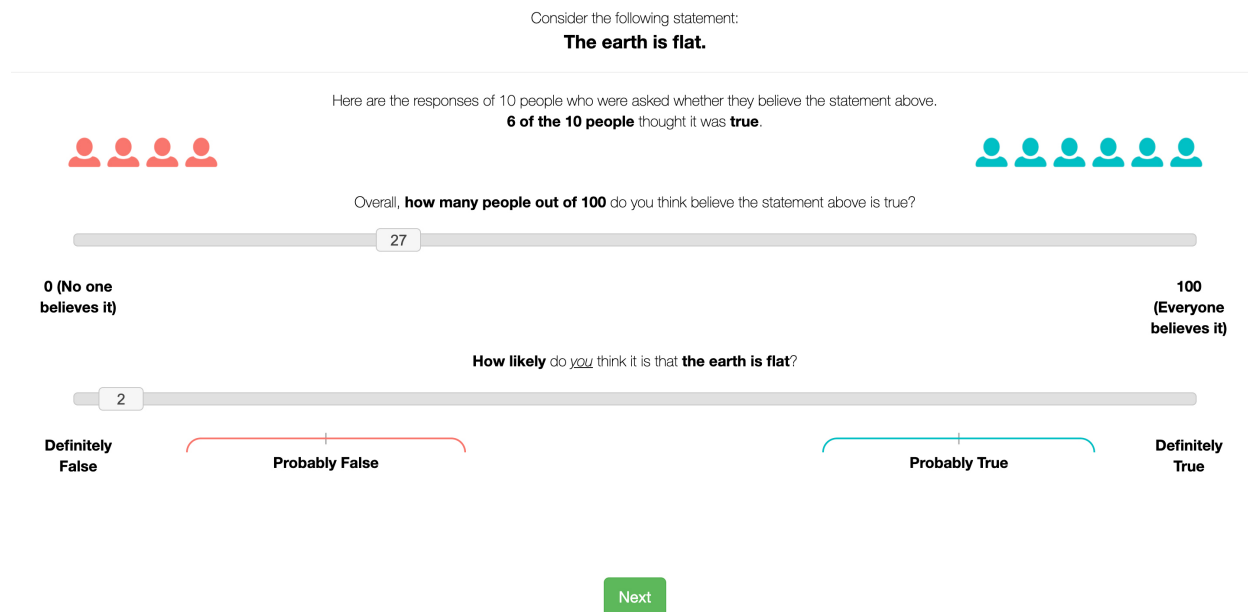


Figure 3.1: Example trial from block 2.

memorize how many of the 10 survey respondents believed each statement in preparation for a memory quiz. Similar levels of belief change across the two experiments would suggest that any observed belief change is not driven by demand effects. Participants' responses in a debriefing survey provide further evidence against the role of demand effects.

Participants

We recruited 608 ($N = 403$ in Experiment 1; 205 in Experiment 2) American adults fluent in English through Prolific (www.prolific.co) to participate in an 18-min online experiment. Only American adults participated because some experimental items referenced U.S.-specific cultural beliefs. We compensated participants at a rate of \$10/hour, with an opportunity to receive a bonus for good performance. We obtained informed consent from all participants and used methods approved by the University of California, Berkeley Institutional Review Board. We preregistered both experiments prior to data collection at <https://aspredicted.org/sy9wv.pdf> (Experiment 1) and <https://aspredicted.org/vc8jr.pdf> (Experiment 2).

3.4 Results

Exclusions

Experiment 1

Eighteen participants were excluded for failing more than one of four attention checks, and an additional 19 participants were excluded for giving blank or unreasonable responses to one or more of six bot-catch questions (e.g., “What’s your favorite frozen treat?”) intermixed between experimental trials. These subject-level exclusions resulted in a final sample of 366.

In addition to these preregistered exclusions, we also excluded trials in which participants initially endorsed the empirically unsupported belief with a likelihood rating of over 60%, because the 40% exaggeration in the Higher Prevalence condition could not be applied to these. These trials constituted 13.1% of our data, and their removal did not significantly affect the interpretation of any analyses.

Experiment 2

Three participants were excluded for failing more than one of four attention checks. No additional participants were excluded, yielding a final sample of 202 participants. Trial-level exclusions (14.6% of all trials) were made using the same criteria as Experiment 1, as preregistered.

Participants revise their estimates of prevalence in light of social data

Experiment 1

If our manipulation worked as intended, participants’ estimates of the prevalence of these beliefs should have increased after seeing the Higher Prevalence data, but remained the same in the Control condition. On average, participants’ prevalence estimates increased by 21.0% in the Higher Prevalence condition and remained relatively stable in the Control condition (decreased by 1.7%). A t-test indicates that this difference is statistically significant ($95\%CI = [22.1, 23.4]$; $t(6904.9) = 65.75, p < .001, d = 1.35$), confirming the effectiveness of the manipulation.

Experiment 2

Participants’ prevalence estimates increased by 25.6% in the Higher Prevalence condition and marginally decreased by 1.0% in the Control condition. A t-test indicates that this difference is statistically significant ($95\%CI = [25.8, 27.5]$; $t(3991.8) = 61.59, p < .001, d = 1.71$).

Participants revise their beliefs in line with social data

Experiment 1

Our main prediction was that participants would increase their endorsement of a belief when given data indicating that the belief is more prevalent than they expect. Figure 3.2 shows the mean amount of belief change per item in each condition for both experiments. As predicted, participants' ratings of the likelihood of these beliefs increased by a mean of 5.44% in the Higher Prevalence condition and remained relatively stable (0.46% increase) in the Control condition ($95\%CI = [4.43, 5.52]$; $t(9046.5) = 17.82, p < .001, d = 0.365$), indicating that exposure to samples with higher belief prevalence influences the plausibility of the belief.

Experiment 2

The main effect of condition was replicated. Participants' likelihood ratings increased by a mean of 6.59% in the Higher Prevalence condition, compared to an increase of 1.22% in the Control condition ($95\%CI = [4.59, 6.15]$; $t(5049.1) = 13.50, p < .001, d = 0.375$).

Belief change is commensurate with change in prevalence estimate

Experiment 1

If social prevalence is treated as an independent source of information that is integrated rationally with prior beliefs, then greater changes in one's estimation of the prevalence of a belief should result in stronger belief updating. That is, a stronger prevalence cue should carry more weight and have a stronger effect on the ultimate belief. Figure 3.3 shows the relationship between change in prevalence estimate and resulting belief change within the Higher Prevalence condition of both experiments. We ran a linear mixed-effects model predicting belief change with condition and change in prevalence estimate as fixed effects and random intercepts per participant and item. This model revealed significant main effects of condition ($\beta = 1.29, t = 3.91, p < .001$) and change in prevalence estimate ($\beta = 0.16, t = 9.01, p < .001$), suggesting that larger increases in prevalence estimates of a belief led to larger increases in personal belief endorsement. There was no significant interaction ($p = .81$). The model accounted for 19.0% of the variance in belief change (conditional R^2).

Experiment 2

A linear mixed-effects model with an identical structure revealed only a significant main effect of change in prevalence estimate ($\beta = 0.16, t = 6.05, p < .001$). Unlike in Experiment 1, the main effect of condition was not significant after controlling for change in prevalence estimate ($p = .16$). The model accounted for 16.4% of the variance in belief change.

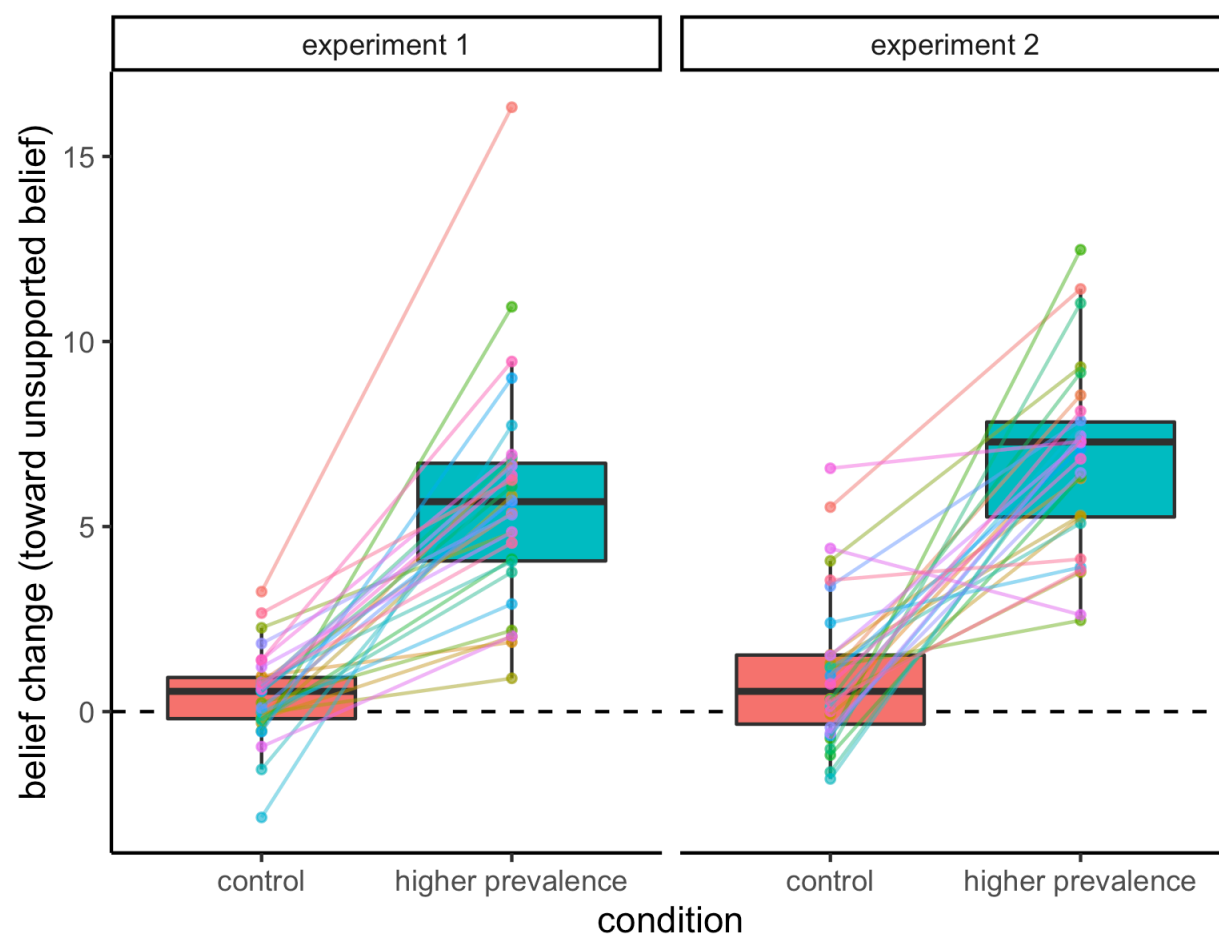


Figure 3.2: Boxplot of belief change after seeing a sample of data that either matched participants' expectations (Control) or indicated a higher prevalence of belief (Higher Prevalence) for Experiments 1 and 2. Points represent the mean belief change per item, with lines showing the effect of condition for each item.

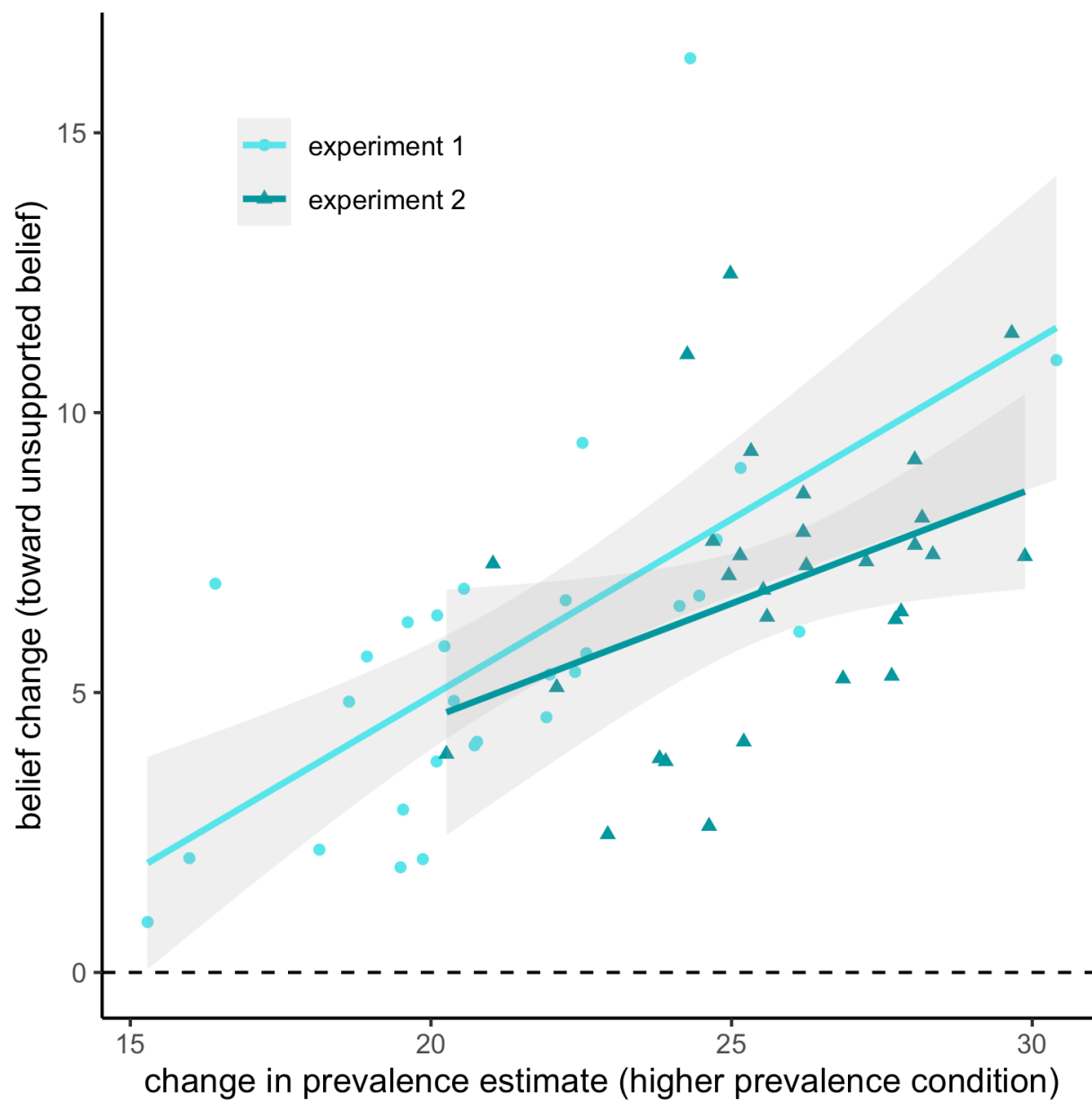


Figure 3.3: Mean belief change vs. mean change in prevalence estimate for each item in the Higher Prevalence conditions of Experiments 1 and 2. Lines represent the fit from linear regression for each dataset with 95% confidence intervals.

Belief change depends on initial certainty

Experiment 1

We predicted that rational belief updating should also depend directly on the initial certainty of the belief. While novel evidence about social prevalence should bear significant weight under conditions of uncertainty, high-certainty beliefs should be relatively resistant to change regardless of social prevalence. Figure 3.4 illustrates the relationship between initial belief (directly related to certainty) and belief change across prevalence conditions. We operationalize certainty as the absolute distance from 50% on the likelihood scale. We fit a linear mixed-effects model with standardized certainty and prevalence condition as fixed effects and random intercepts per participant and item. The model revealed main effects of both scaled certainty ($\beta = 2.42, t = 11.91, p < .001$) and prevalence condition ($\beta = 4.93, t = 18.91, p < .001$), as well as a significant interaction ($\beta = -2.84, t = -10.76, p < .001$). The model accounted for 18.1% of the variance in belief change. Only in the Higher Prevalence condition, where belief change was motivated by data, lower levels of certainty predicted higher belief change as hypothesized.

We replicated our predicted effect of certainty with an additional linear mixed-effects model using change in prevalence estimate as a continuous predictor instead of the dichotomous condition variable. This model had an otherwise identical structure to the first. Again, main effects of standardized certainty ($\beta = 1.33, t = 8.69, p < .001$) and change in prevalence estimate ($\beta = 0.17, t = 25.59, p < .001$) were significant. Crucially, the same significant interaction was observed ($\beta = -1.92, t = -14.99, p < .001$), such that with higher changes in prevalence estimates, initial certainty of a belief negatively predicted belief change. This model accounted for 23.7% of the variance in belief change.

The positive relationship between certainty and belief change in the Control condition is not predicted by rational accounts of belief change and likely stems from an inherent property of our experimental design. Participants' estimates of prevalence were more likely to differ more from their own initial beliefs when their own initial beliefs were more certain. This pattern is consistent with weak priors resulting in a person reverting to their knowledge of social prevalence when making up their mind. Therefore, the difference between the data shown and the participant's own initial belief was higher under conditions of high certainty. To control for this confound, we added the raw difference between the participant's initial prevalence estimate and their own initial belief as a fixed effect to the previous model, and used it to predict belief change in the Control condition. This model revealed that, after controlling for the confound, the unpredicted positive effect of certainty in the Control condition disappeared ($p = .28$). The predicted effects remained: prevalence change ($\beta = 0.21, t = 13.15, p < .001$), the difference between initial prevalence estimate and initial belief ($\beta = 3.85, t = 19.22, p < .001$), and the negative interaction between certainty and prevalence change ($\beta = -1.92, t = -12.18, p < .001$) all stayed significant. The model accounted for 25.7% of the variance in belief change. The negative interaction suggests that certainty also drove belief change in a rational manner in the Control condition when changes

in prevalence estimates were sufficiently high.

Experiment 2

Identical linear mixed-effects models replicated all of the effects from Experiment 1. The first model using prevalence condition as a predictor revealed a significant main effect of scaled certainty ($\beta = 2.48, t = 8.59, p < .001$), a significant main effect of prevalence condition ($/\beta = 5.45, t = 14.58, p < .001$), and a significant negative interaction ($/\beta = -3.75, t = -9.90, p < .001$). The model accounted for 16.5% of the variance in belief change. The second model using prevalence change as a predictor revealed a significant main effect of scaled certainty ($/\beta = 0.87, t = 4.09, p < .001$), a significant main effect of prevalence change ($/\beta = 0.19, t = 19.37, p < .001$), and a significant negative interaction ($/\beta = -1.96, t = -10.33, p < .001$). This model accounted for 18.9% of belief change. As before, after controlling for the confound, the unpredicted positive effect of scaled certainty became nonsignificant ($p = .25$). Also as before, predicted effects in the model still replicated: we observed a main effect of prevalence change ($/\beta = 0.22, t = 9.03, p < .001$), a main effect of the difference between initial prevalence estimate and initial belief ($/\beta = 3.65, t = 11.89, p < .001$), and a negative certainty by prevalence change interaction ($/\beta = -1.02, t = -4.21, p < .001$). This model accounted for 18.6% of belief change.

3.5 General Discussion

In two experiments, we demonstrated that increasing people’s perceptions of the general prevalence of a belief can cause them to endorse that belief more strongly, devoid of any direct evidence. Participants were presented with new prevalence data from an anonymous 10-person sample that conflicted with their prior about a belief’s prevalence by a uniform amount. This social prevalence clue alone was enough to inspire belief change, despite the fact that all items pertained to uncommon, empirically unsupported beliefs from a variety of real-world domains. Additionally, belief change conformed to two key predictions of a Bayesian cue integration framework. First, belief change was sensitive to the reliability of the new prevalence cue. When the new prevalence data was deemed more reliable, as indexed by larger changes in prevalence estimates, belief change increased. Second, belief change was governed by the initial certainty of the participant’s belief, with more weakly held beliefs changing more dramatically. These results were replicated in Experiment 2, thus ruling out the role of demand characteristics. Taken together, these findings suggest that decontextualized, nonauthority social prevalence information serves as a cue that people rationally integrate with existing private evidence according to the relative reliability of both information sources.

It is important to note that participants in our experiment were not simply blindly updating their beliefs to match the prevalence information that they were shown. Recall that

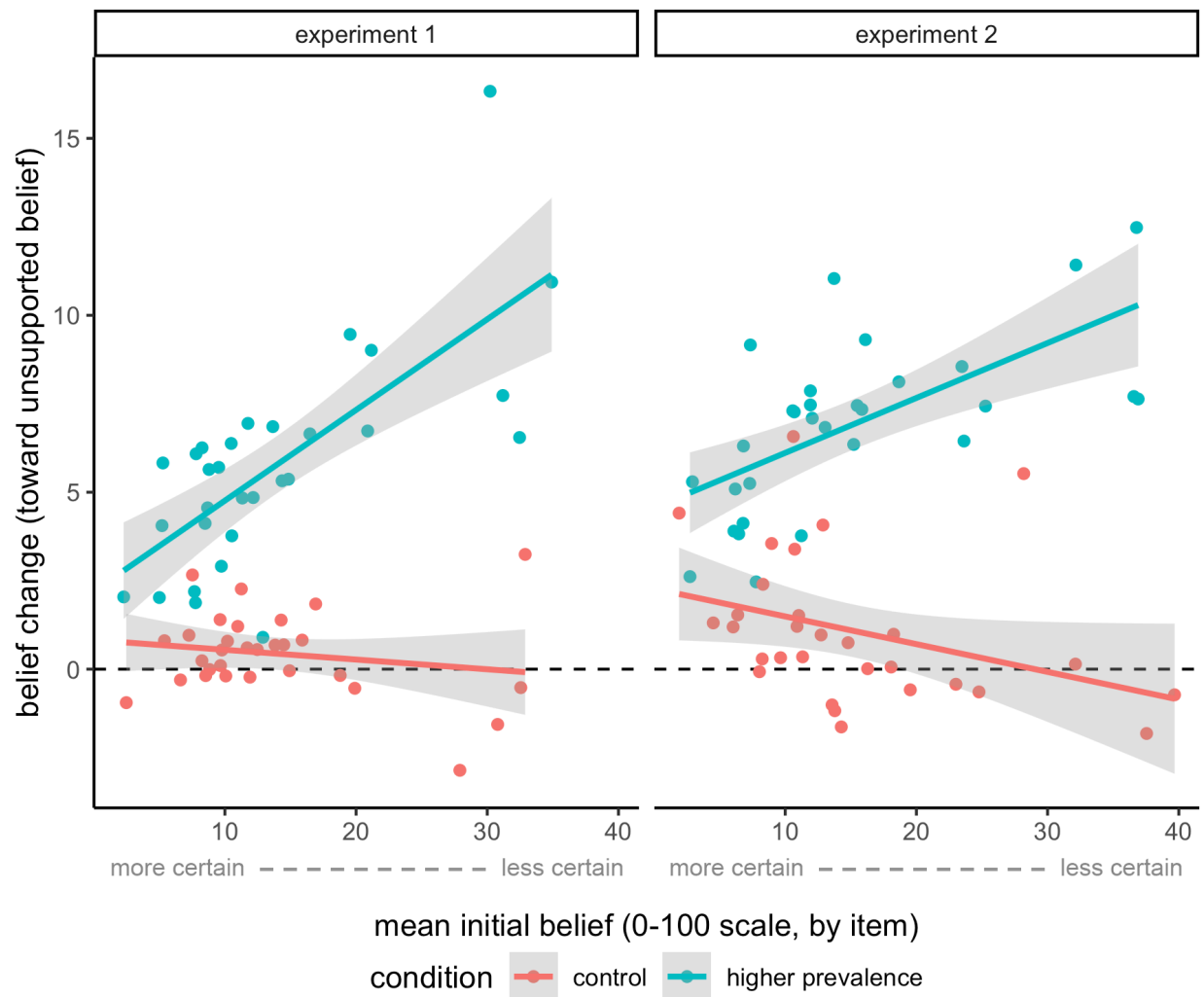


Figure 3.4: Mean belief change vs. mean initial belief for each item, by condition. Lines represent the fit from linear regression for each condition, with 95% confidence intervals. As initial belief increases toward 50% likelihood, certainty decreases.

the prevalence data shown in the Control condition matched participants' initial estimates of the broader prevalence of a belief, and not their personal ratings of its likelihood. Participants were aware of this distinction; their initial prevalence estimates differed from their own initial likelihood estimates by a mean of 16.5% across experiments. Thus, before encountering the prevalence manipulation, participants demonstrated an implicit understanding that their belief may not be representative of that of the broader population. Further, prevalence data in the Higher Prevalence condition was 40% higher than the participant expected, yet participants' final prevalence estimates changed by a mean of only 22.6%. Thus, participants did not uncritically trust the prevalence data, but rather integrated it with their existing belief about the prevalence of each claim.

Further evidence of this rational integration is that changes in prevalence estimates did not always lead to changes in private endorsement of the belief. For example, in Figure 3.4, we see that people do not update their beliefs when they are held with high certainty. A cue integration framework also predicts cases in which consensus information does not motivate belief change. Instead, the prevalence cue is evaluated against the strength of one's prior belief. The possibility of bias or error in social evidence may be particularly salient in domains with widespread ignorance or strong political polarization (Lees & Cikara, 2021). There is evidence that people are able to correct for such bias. Learners rationally discount the weight of a group's testimony when their source is shared and thus statistically dependent (Whalen et al., 2018; but cf. Yousif et al., 2019). Sensitivity to possible selection bias also can lead people to paradoxically draw weaker conclusions from stronger evidence when the evidence is presented by a social agent (Perfors et al., 2018). Future work should investigate cues people use to judge the reliability of prevalence data in paradigms like ours.

Implications for countering false beliefs online

Our findings have direct implications for interventions against the spread of misinformation online. We found that inferring a belief is more prevalent increases the likelihood of belief adoption; thus, high-engagement metrics attached to viral misinformation posts may promote stronger adoption of misinformed beliefs online. This could explain why people rate news from low-credibility sources as higher quality when engagement metrics are present (Chung, 2017). Further, high-engagement metrics elicit more sharing and less fact-checking from users in a simulated social media feed (Avram et al., 2020). This is of particular concern for viral misinformation posts, which can bear engagement numbers in the hundreds or thousands, as opposed to our experimental manipulation with data on 10 people. Hiding social engagement metrics entirely for posts relaying false or misleading information may therefore help reduce false belief. Future work should test interventions along these lines in an effort to counter the online misinformation crisis.

Chapter 4

Young adults are sensitive to how they learn under information overload

(Publication forthcoming: *Shepherd, S. S., *Orticio, E., Castañeda Barba, M., & Kidd, C. (2026). The Brain Rot Effect: The case for adaptive attentional bandwidth. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 48)

4.1 Abstract

Short-form "brain rot" content on platforms like TikTok is widely viewed as harmful for attention and learning. We find that brain rot can benefit learners with higher stimulation needs. We eyetracked young adults ($N = 75$) while they watched physics videos presented alone or alongside a task-irrelevant "brain rot-style" animated video. Individuals who reported preferring brain rot learned better from the brain rot condition than from the blander control condition. Furthermore, eye blinks and pupil dilation suggest that those who prefer brain rot were less overstimulated by brain rot content and required more effortful executive control to sustain attention to control content. Finally, preference for brain rot in our task is associated with higher media consumption and tolerance for high stimulation. These findings may point to individual variation in optimal cognitive load and provide preliminary support for an adaptive account of attentional bandwidth.

4.2 Introduction

Trends in digital media favor fast paced, visually stimulating features to capture attention (Cutting, 2016; Simon et al., 1971). The emergent attention economy has given rise to extreme formats like "brain rot", where two unrelated videos are presented simultaneously. Although brain rot content might sound as though it should introduce distraction into learning, its popularity might suggest otherwise. We use brain rot to test whether attention adapts to increased perceptual load.

Features of brain rot are viewed as harmful

Both popular discourse and academic research link short-form and high-stimulation media trends to shorter attention spans, digital addiction, diminished critical thinking, and attentional disorders (Christakis et al., 2018; Hayes, 2025; Lissak, 2018; Nguyen et al., 2025; Ward et al., 2017; Wu, 2016). Leading theories suggest the heightened audiovisual stimulation disrupts processing and learning (Christakis et al., 2012; Lillard & Peterson, 2011; Nguyen et al., 2025; Shepherd & Kidd, 2024).

Then why is brain rot popular?

Brain rot was the the Oxford word of the year in 2024, and refers to simultaneous video streams designed to capture and hold attention on social media platforms like Instagram and Tik Tok (Heaton, 2024). Brain rot can also refer to the assumed cognitive decline due to overconsumption of this type of content (Yousef et al., 2025). Here, we operationalize brain rot as two videos presented simultaneously via split screen, with a primary video alongside a second visually engaging but irrelevant video.

The recent appearance of study tools that convert students' notes into brain rot suggests that at least some people believe this content facilitates learning (Upton-Clark, 2024). Brain rot could theoretically improve processing for individuals who need greater stimulation for optimal learning, in line work on learning from accelerated playback speeds (Chen et al., 2024; Lang et al., 2020; Tharumalingam et al., 2025). We test the idea that brain rot may not uniformly impair processing and downstream learning for all learners, and could instead provide greater stimulation for learners who need more, thus facilitating their learning.

Distinct sources of cognitive load

Human attention is limited; thus, the system must allocate resources to support effective perception, processing, and learning with respect to resource limitations. Cognitive load refers to the demands that a task places on these limited resources. The specific features of a task impact cognitive load. Perceptual complexity—like a busy visual scene—increases the need to focus on task-relevant information and ignore task-irrelevant distractors. On the other hand, executive control requires resources to stay on task and hold relevant information in working memory, among other executive functions (Kahneman, 1973; Reese et al., 2016; Sweller et al., 1998). Current models of processing distinguish between perceptual load and executive load.

The load theory of attention provides a framework that accounts for the distinct effects of perceptual and executive load on task performance (Lavie et al., 2004). Task performance may improve under higher perceptual load, counter to traditional accounts of cognitive load. These findings are explained by a two-stage account. Under high perceptual load, an early perceptual mechanism filters out distractors before they are processed. Under low perceptual load, distractors are processed and must be suppressed by a late-stage executive control

mechanism. As a result, high perceptual load reduces distractor interference, while high executive control load increases interference, a pattern widely supported by both theoretical and empirical work (Lavie, 2005, 2010).

These differential effects demonstrate that processing is not only sensitive to the degree of cognitive load, but also to the source of load. In this view, people may seek higher perceptual load (like brain rot) to maintain a balance between perceptual and executive load. Variation in preferences for highly stimulating media, as well as mixed findings in the playback-speed literature, may be explained by variation in this optimal range of stimulation.

Brain rot as an adaptive attentional strategy

Here, we investigate brain rot as a potential case study of how attentional capacity may be adaptive rather than fixed. Individuals may vary in the range of perceptual stimulation that optimally supports engagement and processing. Too much overstimulates and impairs processing, resulting in reduced performance, while too little under-stimulates and requires greater executive effort, also reducing performance. If true, this would predict that high perceptual load would not uniformly impair learning. Rather, the impact would depend on the (mis)match between the properties of the stimulus and the individual's optimal range of stimulation.

We test this account by manipulating perceptual load with brain rot stimuli during a learning task. We find that the same material disrupts processing for some participants while improving processing for others, and that the effect is predicted by their post-test preference for brain rot. Exploratory eye-tracking analyses further support an adaptive attentional bandwidth framework by differentiating perceptual and executive load.

4.3 Methods

Participants

We recruited seventy-five undergraduate psychology students ($M_{age} = 20.8$, $SD = 2.7$, range = 18-35) from a university subject pool. We excluded fourteen additional participants due to technical issues ($n = 9$), excessive environmental noise ($n = 4$), or experimenter error ($n = 1$). We compensated participants with course credit and a \$5 bonus.

Stimuli

Primary stimuli were two excerpts of educational physics videos lasting 2 minutes and 15 seconds.¹ The physics videos were presented on the bottom half of a split-screen visual display (see Figure 4.1). The top half consisted of either a static gray background or an

¹Adapted from 'Windmills Are NOT Like Dams' and 'How Modern Light Bulbs Work' by @minutephysics on YouTube.

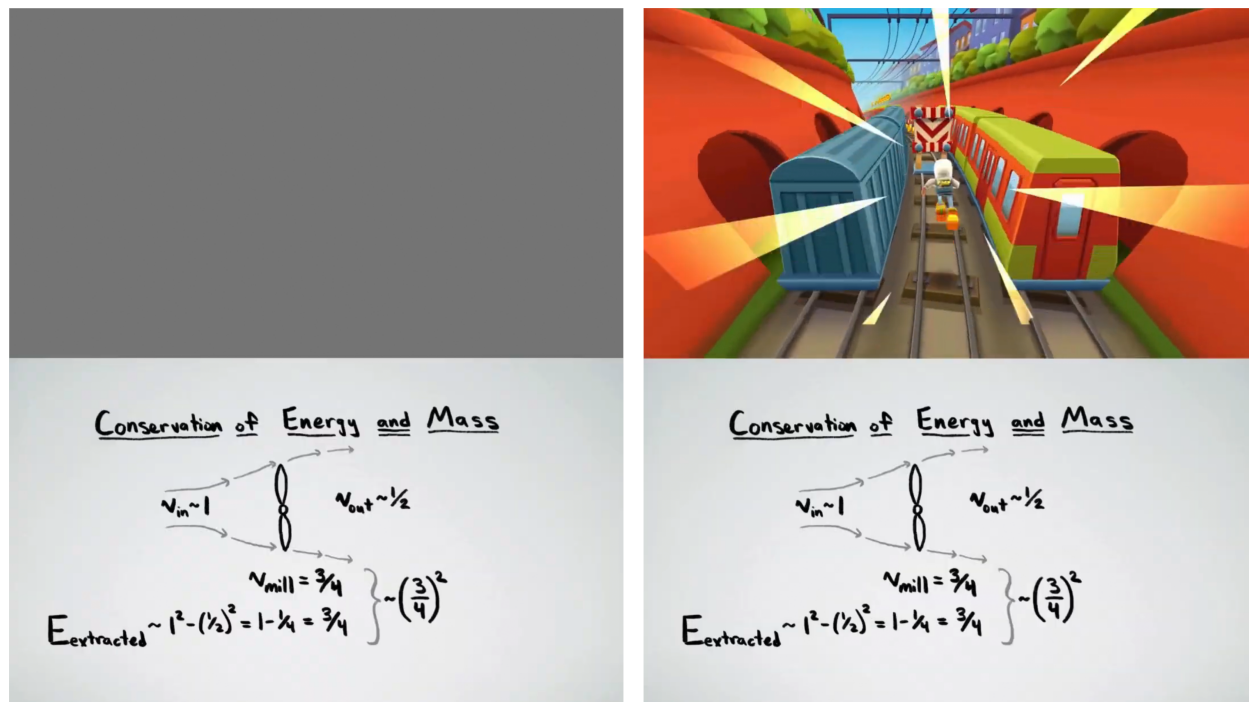


Figure 4.1: Screenshots of matched videos for control condition (left) and brain rot condition (right).

unrelated, visually engaging gameplay video², according to a counterbalanced, within-subject trial condition (control vs. brain rot). We paired each video with a 30-item multiple-choice test assessing learning of the physics material.

Procedure

Participants completed the study in a quiet, darkened room to control for distractions. We presented the visual stimuli on a monitor connected to a desktop-mounted EyeLink 1000 eye tracker (SR Research Ltd. 2005). The EyeLink monocularly recorded gaze location and pupil diameter at 1000Hz. We positioned participants 70 cm from the display. We calibrated the eye tracker using a standard nine-point calibration procedure followed by an identical validation sequence before each trial. We repeated the validation sequences until satisfactory, < 0.5 degrees of average error.

Participants watched a video and completed a corresponding comprehension test in each of two within-subject blocks. Participants were required to view each video in its entirety at least once, but were allowed to re-watch the video continuously until they felt ready to

²Adapted from Subway Surfers Gameplay [4K 16x9] No Copyright by @subwaysurfers on YouTube.

complete the test. After the initial viewing, the video looped automatically until participants looked away from the screen for more than one second. Following each video, participants completed a 30-item test assessing comprehension of the video content. Following both blocks, participants gave their preferences for the two video presentations and answered questions about their media consumption habits. We counterbalanced video content across participants and randomized trial condition across trials.

Brain rot preference score

We created a composite score of brain rot preference from five equally weighted, binary questions. We asked participants which of the two videos they found more engaging, more enjoyable, more effective for learning, easier for sustaining their attention, and more preferred overall. Higher composite scores indicate a stronger preference for the brain rot video presentation over the control presentation.

Media habit score

We created a media habit score to quantify participants' prior consumption of online media. This composite score was composed of 9 range-normalized and equally weighted questions, including 1 binary and 8 ordinal items. These items asked participants about their screen time, time spent watching short-form content, frequency of watching sped-up content, frequency of watching brain rot content, frequency of consuming multiple streams of content simultaneously, frequency of consuming unrelated media while working or studying, attitude toward engagement with brain rot content, experience using brain rot study tools, and interest in using brain rot study tools. Higher composite scores correspond to heavier online media consumption, stronger proclivities for highly stimulating content, and stronger interest in educational applications of brain rot.

Eye-tracking data pre-processing

Pupil dilation

We used pupil size as a proxy measurement of executive load. To quantify pupil size, we removed 150ms before and after a blink event to account for blink-induced pupillary response (Bentivoglio et al., 1997). We then normalized pupil size for each participant using the first 10 seconds of the current trial to account for different individual baselines for each trial type (brain rot and control). Finally, we scaled pupil change from baseline for clearer interpretation.

Eye blink behavior

Endogenous eye blink behavior reflects the regulation of sensory and information processing (for review see Callara et al., 2023; Stern et al., 1984). We analyzed blink behavior to test

the effect of increased perceptual load in the brain rot condition as a function of participants' preference scores using two metrics: total blink time and blink count.

Total blink time is the total number of eye-tracking samples occurring during a blink event. Since blink event samples and missing data cannot easily be differentiated, all blink events longer than one second were removed, a standard threshold for identifying endogenous eye blinks (Stern et al., 1984). Participants used a chin rest, so any blink events longer than one-second are likely due to drowsiness rather than missing data. This data filtering resulted in the removal of about 0.6% of all data. The final blink sample count was summed and divided by remaining samples. The final measure is total blink time per sample for a given trial. Blink count was also quantified as distinct blink events. Both total blink time and blink count were log-transformed prior to analysis to meet model assumptions (Bentivoglio et al., 1997).

4.4 Results

Brain rot has heterogeneous effects on learning and metacognition

We assessed whether the simultaneous, brain rot-style presentation of unrelated video game content during viewing of physics videos affected performance on a subsequent comprehension test. We ran a generalized linear mixed-effects model predicting trial-level accuracy, with condition (control vs. brain rot) and condition order as fixed effects and by-participant random intercepts and condition slopes. This model revealed a significant effect of condition order ($\beta = 0.29$, $t = 3.35$, $p < .001$) such that performance was better in the second within-subjects condition. There was no significant effect of condition ($\beta = -0.14$, $t = -1.56$, $p = .12$). There was no group-level effect of brain rot on test performance. However, this null effect on the group-level obscured substantial heterogeneity in the effect of brain rot across participants. Consistent with this variation, the mean absolute difference in by-participant average accuracy between conditions was 15.5% (SD = 0.12, range = 0-0.5). Our sample was evenly split between participants who learned better in the control condition (49.3%) vs. the brain rot condition (45.3%).

We also tested for condition differences in participants' metacognitive sensitivity, or the relationship between certainty and accuracy. We quantified metacognitive sensitivity using the pROC package in R to calculate the area under the Type 2 ROC curve, or AUC2. This curve captures the hit rate vs. false alarm rate for whether higher trial-level certainty predicts accuracy. An area of 0.5 corresponds to at-chance performance, and higher AUC2 values indicate stronger metacognitive sensitivity.

A linear mixed-effects model predicting AUC2, with condition (control vs. brain rot) and condition order as fixed effects and random intercepts per participant, revealed no significant effect of condition ($\beta = 0.03$, $t = 1.22$, $p = .23$) or of condition order ($\beta = 0.03$, $t = 1.50$, $p = .14$). The null effect of condition reflects bidirectional effects on the individual level. The mean by-participant difference in AUC2 between conditions was 15.1% (SD = 0.13,

range = 0.003-0.6). Our sample was again evenly split between participants who had higher metacognitive sensitivity in the control condition (49.3%) vs. the brain rot condition (49.3%). For most participants (65.3%), the condition they learned better in was also the condition in which they showed better metacognitive sensitivity.

Brain rot preferers learn better from brain rot, worse from traditional content

We used participants' reported preferences for the brain rot-style video presentation or the control presentation to test whether participants were aware of which condition was more conducive to their learning. Brain rot preference scores were right-skewed ($M = 1.49$ out of 5, $SD = 1.79$). Nearly half ($n = 35$, or 46.7%) of the participants had a score of zero, meaning that they preferred the control presentation across all five questionnaire items. Eight (or 10.7%) participants had a brain rot preference score of 5, reflecting a universal preference for the brain rot presentation. Those who learned better in the brain rot condition had higher brain rot preference scores ($M = 2.03$) than those who learned better in the control condition ($M = 0.68$), indicating that subjective preference was calibrated to learning outcomes.

We assessed whether participants' post-test content preferences predicted different learning outcomes, and found a crossover effect (see Figure 4.2). Participants with a stronger brain rot preference performed better in the brain rot condition, whereas participants with a weaker brain rot preference performed better in the control condition. We fit a linear mixed-effects model predicting participants' average accuracy, with condition (control vs. brain rot), scaled brain rot preference score, and condition order as fixed effects and random intercepts per participant. This model revealed a significant effect of condition order ($\beta = 0.04$, $t = 2.08$, $p = .04$), a significant effect of scaled brain rot preference score ($\beta = -0.05$, $t = 2.55$, $p = .01$), and a significant interaction between condition and scaled brain rot preference score ($\beta = 0.08$, $t = 3.59$, $p < .001$). The effect of the brain rot condition was not significant ($\beta = -0.02$, $t = -1.22$, $p = .22$). Higher brain rot preference scores predicted lower test performance in the control condition, but higher test performance in the brain rot condition.

Brain rot preferers have poor metacognition for traditional content

Stronger brain rot preference predicted worse metacognitive sensitivity in the control condition and better metacognitive sensitivity in the brain rot condition (see Figure 4.3). We fit a linear mixed-effects model predicting AUC2, with condition (control vs. brain rot), scaled brain rot preference score, scaled average accuracy, and condition order as fixed effects and random intercepts per participant. The model revealed a significant effect of brain rot preference score ($\beta = -0.04$, $t = -3.02$, $p = .003$), a significant effect of scaled average accuracy ($\beta = 0.06$, $t = 6.42$, $p < .001$), and a significant interaction between condition and brain

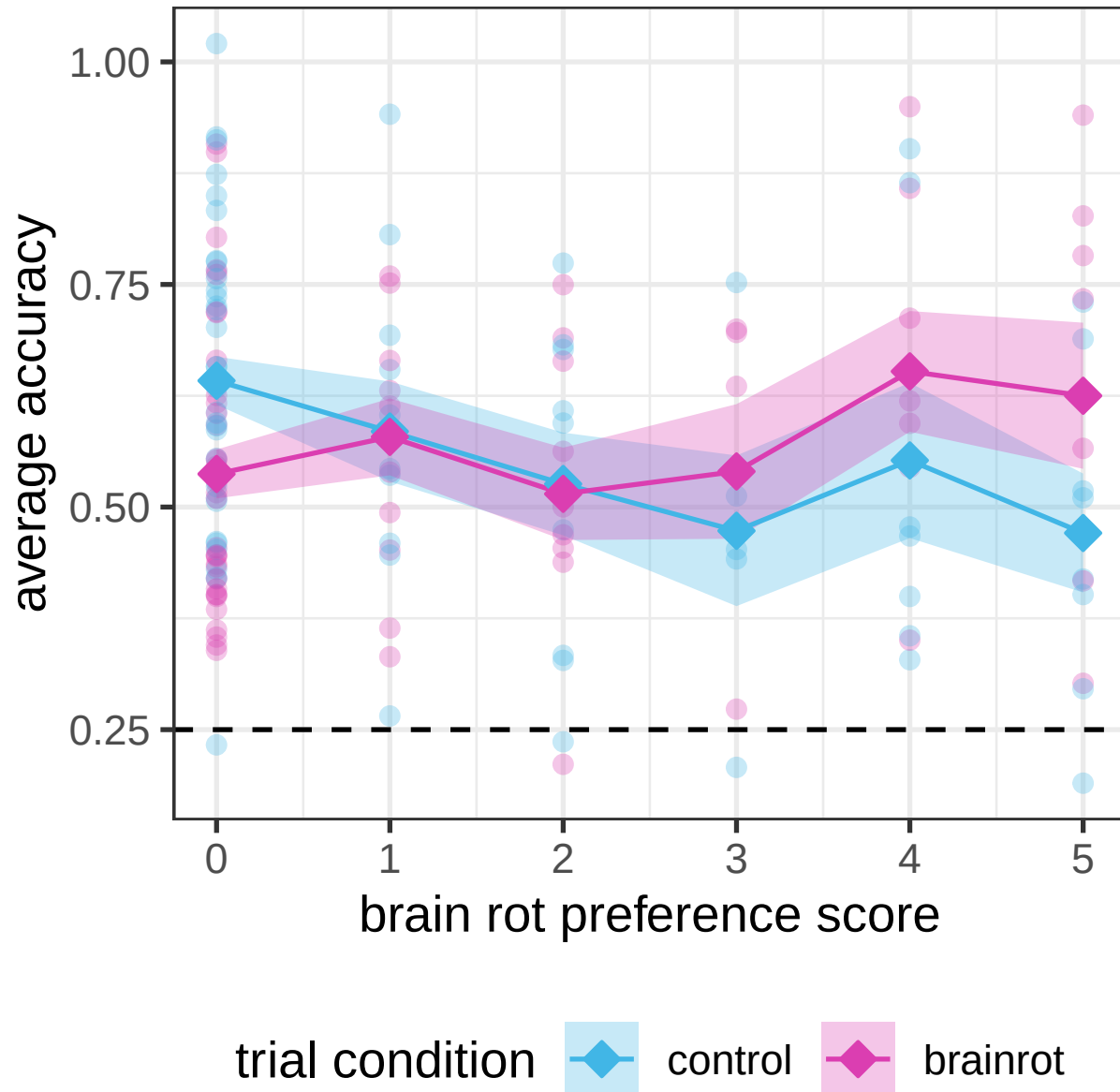


Figure 4.2: Preference for brain rot content predicts differential comprehension test performance across conditions. Dashed line represents chance performance. Points are participant averages by condition and are jittered vertically. Diamonds are condition means by preference score. Ribbons represent SE.

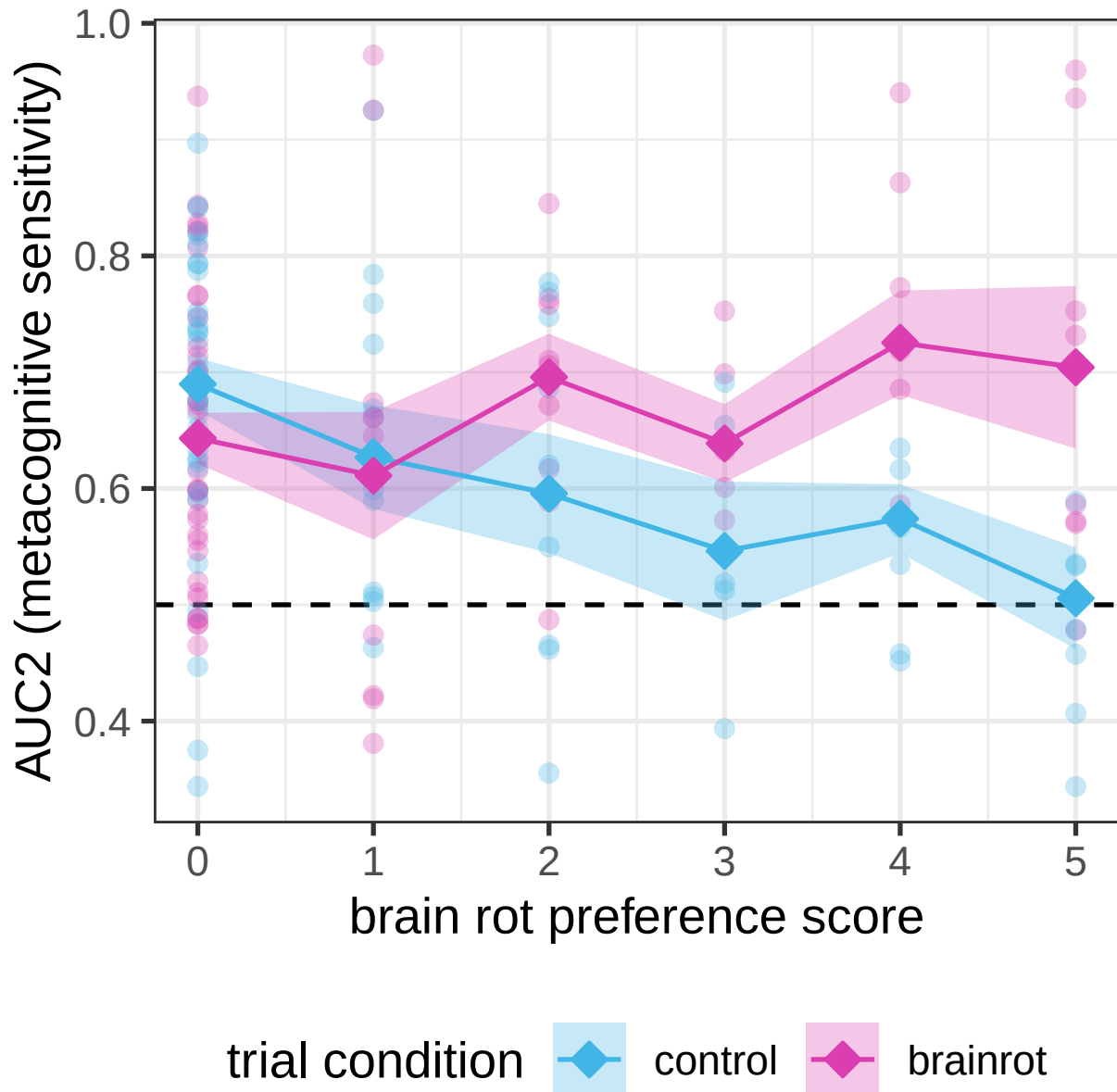


Figure 4.3: Preference for brain rot predicts differential metacognitive sensitivity across conditions. Dashed line represents at-chance metacognition. Points are participant averages by condition and are jittered vertically. Diamonds are condition means by preference score. Ribbons represent SE.

rot preference score ($\beta = 0.06$, $t = 2.97$, $p = .004$). Better task performance was associated with higher AUC2 scores, as is common with many measures of metacognition (Rahnev, 2025). Critically, the interaction between condition and brain rot preference remained after controlling for task performance. Higher brain rot preference was associated with enhanced metacognition in the brain rot condition, but diminished metacognition when learning from the less stimulating control presentation.

Brain rot preference is not captured by gaze patterns

Looking behavior was analyzed within the brain rot condition alone since looking to the video was at ceiling during the control condition. A linear regression predicting the proportion of time spent looking to the content video during brain rot trials (as opposed to the brain rot video) revealed no effect of brain rot preference score ($\beta = -0.02$, $t = -1.16$, $p = 0.25$) or condition order ($\beta = 0.03$, $t = 0.85$, $p = 0.4$). A linear regression predicting the number of switches made between the brain rot and content videos similarly showed no significant effect of brain rot preference score ($\beta = 6.4$, $t = 1.16$, $p = 0.25$) or condition order ($\beta = -19.24$, $t = -1.7$, $p = 0.09$). These results indicate that preference for brain rot does not predict differences in looking behavior in our task.

Lastly, the duration of watching did not differ between conditions or preference scores. A linear mixed-effects model predicting total viewing time confirmed no effects of condition ($\beta = .018$, $t = 0.12$, $p = 0.9$), brain rot preference score ($\beta = 0.07$, $t = 0.58$, $p = 0.57$), or condition order ($\beta = -0.2$, $t = -1.26$, $p = 0.21$), and no interaction between trial condition and brain rot preference score ($\beta = 0.05$, $t = 0.37$, $p = 0.71$). Since preference score does not predict watching duration in the brain rot condition ($M = 2.46$, $SD = 1.08$) or the control condition ($M = 2.51$, $SD = 1.1$) we restrict our analysis of eye blink and pupil dilation to the first loop of each video.

Eye blink behavior supports variation in experience of perceptual load

A linear mixed-effects model predicting the total blink time (log-transformed blink samples) with condition, scaled brain rot preference score, and condition order as fixed effects and random intercepts per participant, revealed a significant interaction between condition and preference score, ($\beta = -26.55$, $t = -2.28$, $p = .025$). This suggests that only individuals with lower preference for brain rot experience greater perceptual load while watching brain rot, reflected by their increased total blink time. The effect of preference score was significant in the brain rot condition ($\beta = -26.271$, $t = -2.30$, $p = .02$), but not in the control condition ($\beta = 0.27$, $t = .024$, $p = .98$). Total blink time is reduced among brain rot preferers in the brain rot condition only.

We also analyzed distinct eye blink events as opposed to total eye blink durations. A linear mixed-effects model predicting log-transformed blink count confirmed that higher preference results in fewer blinks in the brain rot condition ($\beta = -0.23$, $t = -2.19$, $p = .03$). In

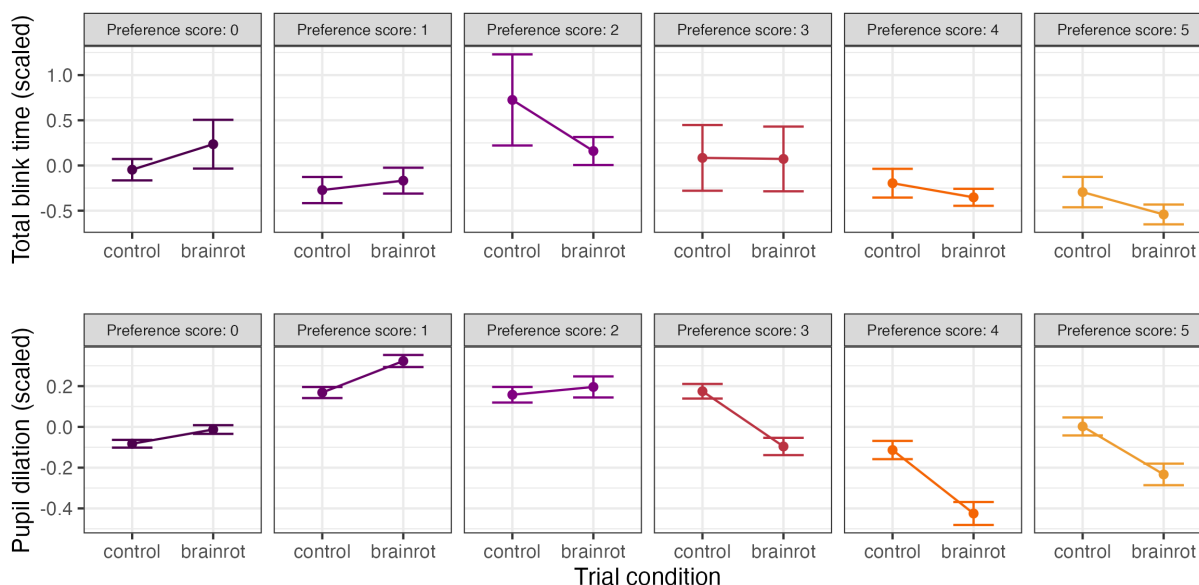


Figure 4.4: Preference for brain rot predicts differential blink and pupil behavior across conditions. Points are average blink time by condition (above). Points are average change in pupil size from baseline by condition (below). Error bars represent SE.

the control condition, preference did not significantly predict blink count ($\beta = -0.11$, $t = -1.08$, $p = .28$). The interaction between distinct blink events and trial condition was not significant for distinct blink events ($\beta = -0.12$, $t = -1.47$, $p = .15$). These effects remain for both blink measures after controlling for task performance and metacognitive sensitivity.

Pupil dilation suggests increased executive load during traditional content for brain rot preferers

We used a linear mixed-effects model to analyze the effects of condition and scaled brain rot preference score on participants' normalized pupil size, controlling for condition order with a random intercept for participants. The main effects of condition ($\beta = -3.66$, $t = -1.04$, $p = 0.29$) and of brain rot preference score ($\beta = 7.77$, $t = 0.45$, $p = 0.66$) were not significant. Condition order significantly predicted pupil dilation ($\beta = 21.64$, $t = 5.14$, $p < 0.001$). A significant interaction between brain rot preference score and condition ($\beta = -40.65$, $t = -9.85$, $p < 0.001$) revealed that a stronger preference for brain rot is associated with decreased pupil size while watching brain rot. These results are consistent with decades of empirical work demonstrating that pupil size increases with general cognitive load, as well as more recent evidence that pupil dilation specifically reflects increased task-related executive control demands (Gilzenrat et al., 2010; Kahneman, 1973). In line with this interpretation,

increased pupil dilation in the control condition among brain rot preferers indicates increased executive demand. This suggests that attending to the control condition was more effortful for brain rot preferers.

Brain rot preferers consume more media

We assessed whether participants' digital media habits predict their preferences between the video presentations in our task. Media habit scores were approximately normally distributed ($M = 3.83$ out of 9, $SD = 1.39$, range = 0.90-7.75). Media habit score was positively correlated with brain rot preference score (Spearman's $\rho = 0.35$, $p = .002$). Media habits characterized by heavier online media consumption, higher tolerance for attentional load, and increased interest in or use of brain rot study tools were associated with a higher likelihood of preferring to learn from the brain rot video presentation in our task. Notably, media habit scores did not independently predict task performance or metacognition in subsequent analyses.

4.5 Discussion

The incentives of digital media environments are rapidly changing the nature of the online content that we consume. Attention-grabbing "brain rot" content exemplifies this change. The term brain rot itself codifies a set of lay hypotheses, largely untested, about the potential (harmful) effects of brain rot on attention and learning. Here, we show that, counter to widespread assumptions, brain rot-style presentations of otherwise identical content led to improved comprehension and metacognitive sensitivity for a meaningful subset of learners. These results are in line with hypotheses that suggest that individuals learn best from stimuli that meet their specific optimal loads.

Our data on individual preferences for learning from brain rot suggest that participants are at least somewhat well calibrated in understanding the attentional demands that are optimal for their own learning. In this sense, our composite score of brain rot preference may be seen as a measure of global metacognition for brain rot. Global metacognition represents broad, more abstract self-beliefs about performance. This contrasts with local metacognition, which refers to momentary confidence in trial-by-trial performance and has overlapping but distinct neurocognitive underpinnings (Rouault & Fleming, 2020). Awareness of one's ability to learn from brain rot content is a double-edged sword. The growing use of brain rot even in educational settings may be less concerning if the primary users genuinely benefit. At the same time, a perceived dependence on high stimulation for learning may lead to disengagement from traditional educational content.

Attentional signatures of brain rot learners

Eye-tracking findings highlight two key attentional signatures associated with improved learning from brain rot. Importantly, these signatures are not a function of visual search. Neither proportion of looking time nor frequency of switches to the brain rot predicted whether participants learned better from brain rot. Instead, these performance patterns covary with two key differences in visual processing mapping onto the two-stage load theory of attention (Lavie et al., 2004). Blink responses indicate divergence in perceptual load, while pupil dilation patterns point to variation in executive load.

First, people who learn better from brain rot are less overstimulated by brain rot. More extreme blink responses correspond to higher perceptual load, or overstimulation, and brain rot learners blinked less in the brain rot condition (Bentivoglio et al., 1997). By contrast, people who learn better in the control condition demonstrate stronger signs of overstimulation in the brain rot condition, as expected (see Figure 4.4). Brain rot learners' apparent immunity to this overstimulation is consistent with an attentional system adapted to a higher objective ceiling of perceptual stimulation. Crucially, this variation exists despite the two groups viewing identical content, indicating that perceptual load is not objectively determined by stimulus characteristics alone. Our findings suggest that observer factors also affect perceptual load.

Second, people who learn better from brain rot exhibit higher executive load when watching content *without* brain rot. Pupil dilation indexes executive load (Beatty, 1982; Gilzenrat et al., 2010). We observed a crossover effect (see Figure 4.4) in which participants tended to show higher pupil dilation when watching the video from which they learned less. Brain rot learners showed higher pupil dilation in the control condition, reflecting increased expenditure of executive control to maintain attention on the control video. Speculatively, this control video was too boring for brain rot learners, i.e., outside of their optimal range of stimulation. This is striking given that the videos in the control condition were still very stimulating, with constantly changing visuals corresponding to the concepts being taught. Thus, our task may underestimate the extent of variation in attentional bandwidth within our sample. Simpler content would likely produce stronger effects than we observed here.

Taken together, these findings lend preliminary support to an adaptive attentional bandwidth account. The range of stimulation or load that is optimal for learning may not be fixed, but rather vary across individuals. Individual adaptation shifts, but does not expand, this optimal window. This property is reflected in the performance tradeoffs we observe in our task, where ability to learn from perceptually demanding content is associated with poorer outcomes when learning from less stimulating content. Still, our paradigm relies on pre-existing variation and does not establish adaptation. Future work should use longitudinal and intervention designs to test this account more directly.

Connection to real-world digital media habits

Preliminary evidence suggests that real-world digital media habits relate to preference for learning from brain rot. Specifically, we found that preference for learning from brain rot was associated with a composite score indicative of higher online media consumption, tolerance for high stimulation, and intentions to use brain rot study tools. This correlation does not establish causality, although we cautiously speculate that the relationship is complex and bidirectional. It is likely that increased short-form media consumption breeds familiarity with, and liking of, brain rot content, which in turn reinforce digital habits. Other more stable individual differences may also affect people’s propensity to engage with these forms of content.

Public discourse increasingly emphasizes “resetting” attention and building tolerance to boredom through incremental exposure to “boring” tasks, like reading or even a viral TikTok challenge to do nothing (Nazaryan, 2026; D. G. Smith, 2023). These trends suggest that attention can be adapted in the opposite direction than observed in our brain rot learners, lowering the baseline stimulation needed to maintain engagement. Beyond individual practices, this discourse is part of a broader “attention activism” movement, including initiatives of groups like the School of Radical Attention, where disengagement with content is reframed as a form of resistance to algorithmically-curated feeds and content competing for attention (Burnett & Mitchell, 2025; The Friends of Attention, 2026). The adaptive account that we provide evidence for here would suggest that these movements and practices could be effective.

Limitations

Properties of our experiment and sample warrant caution when generalizing these claims. We used only two physics videos and one gameplay distractor video, so it is unclear whether these effects would replicate with other stimuli or in other domains of learning. In addition, our sample was limited to undergraduate students whose age and media habits are unrepresentative of the broader population. We suspect, for example, that other demographics may differ, both in the proportion of users whose learning would benefit from highly stimulating content and in the malleability of their attentional systems. Familiarity with brain rot content may also be necessary to accurately assess one’s ability to learn from it. Finally, participants were incentivized to learn from the videos in preparation for a comprehension test, so the observed effects may differ in more naturalistic, entertainment-oriented contexts where consumption of brain rot may be more frequent.

4.6 Conclusion

We proposed an adaptive account of attentional bandwidth in which processing is best facilitated within an individual’s optimal range of stimulation. We found preliminary evidence in line with this account using the brain rot trend as a test case. In general, our results

indicate that an individual's optimal range of stimulation may shift, perhaps as a result of a combination of media habits and more stable individual differences.

Chapter 5

Conclusion

5.1 Learning how to learn online

A key theme of this research is that epistemic competencies develop through direct engagement with contemporary information environments. In Chapter 2, children's first-hand encounters with detectable inaccuracies led them to spontaneously increase fact-checking behavior in a digital environment without instruction or explicit training. In Chapter 3, adults adopted questionable beliefs most strongly when their initial beliefs were uncertain, likely because they lacked direct knowledge to anchor their judgments. Finally, in Chapter 4, young adults demonstrated a calibrated sensitivity to the optimal level of stimulation for their learning, which drew upon their first-hand experiences in our task and likely also their lived experiences of consuming attentionally demanding media online. These studies underline the potential for our cognition to adapt to new information environments. Learners need hands-on experience with the environments to which they are adapting.

This insight informs intervention strategies. First, protective strategies that restrict exposure to the complexity of digital media environments may leave learners less equipped to navigate these environments in the future. Second, interventions that are passive or abstracted from their real-world context may have limited effectiveness. This is not an argument for throwing learners into the deep end without supports or abandoning efforts to curate information environments. Rather, this work emphasizes that epistemic competence is built through active engagement with imperfect information. Interventions and educational strategies should create low-stakes, structured opportunities for learners to understand the risks of real information ecosystems and put healthy epistemic practices into action.

Crucially, though, real-world behavior in digital environments is messy. The tasks in this research approximate aspects of digital experiences, but the experimental context limits ecological validity by drawing attention to learning goals. In more naturalistic contexts, users are motivated by other competing goals, like entertainment and social connection. It is unclear to what extent these competencies would emerge during more typical, passive digital consumption. Moreover, the gap between epistemic competence and epistemic action

can be vast. It is one thing to be aware that a claim is dubious, or that observed data are unrepresentative. Translating that awareness into action is another thing entirely. This often involves complex skills like identifying and evaluating relevant counterevidence or reasoning about the opaque generative process of algorithm-provided data. Future work is needed to understand the starting point and developmental trajectories of these critical skills, especially in ecologically valid settings.

5.2 Epistemic environmentalism and the importance of trust

While healthy skepticism is important, it is perhaps even more vital to preserve calibrated trust. In many cases, being uninformed may be more common than being misinformed (Allen et al., 2020). Thus, hyperskepticism can be just as harmful as credulity. Generalized distrust makes people vulnerable to conspiracy beliefs (Thielmann & Hilbig, 2023), which can breed further distrust. This is a concrete concern given the eroding trust in science, institutions, and media across the world. Critically, the structure of digital information ecosystems exacerbates epistemic risks. The sheer volume of information available online can be overwhelming. When people lose confidence in their ability to sift through this information, they can become apathetic and disengage (Rippon, 2024). Bad actors can exploit this vulnerability by intentionally "flooding" the information environment with noisy, incongruent information (Anderau, 2023). The result of this disinformation strategy is that learners are overloaded and less trusting of reliable information (Szegofi et al., 2026). Misinformation doesn't need to persuade, only to confuse. To counter this, we need to take long-term measures to restore faith in our information ecosystems.

Epistemic environmentalism (see Ryan, 2018) may be a helpful framework for intervention. First, we should all be mindful of our epistemic practices. Our individual capacities for adaptive information-seeking and belief formation are meaningful, and this dissertation has sought to understand them. At the same time, they operate within an environment shaped by many structural forces, like recommendation algorithms, platform incentives, and regulatory frameworks. The power to change these forces is heavily concentrated between a small number of extremely powerful technology companies. A crucial part of this movement, as in geo-environmentalism, is to avoid the individualization of responsibility. An outsized emphasis on individual-level strategies risks deflecting accountability from larger, systemic drivers (Anjum & Aziz, 2026). Technology platforms and regulatory institutions, as well as individuals, are responsible for reducing epistemic pollution and improving the epistemic environment for the next generations of learners.

5.3 Summary

In sum, this dissertation contributes to an ecological theory of belief formation in digital environments. It examines key epistemic challenges posed by contemporary information environments, particularly social media, and identifies competencies that enable learners to adapt to these challenges. These competencies are attuned to the structural features of digital environments and responsive to first-hand experience. Importantly, these competencies may be capable of development. Thus, this work offers a foundation for understanding how healthy digital epistemic practices can be cultivated across the lifespan. Improving access to truth requires an analysis of the alignment between our cognition and the information environments we inhabit, both online and offline.

Bibliography

- Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? children consider informants' sources when deciding whom to believe. *Journal of experimental psychology: general*, 151(10), 2481.
- Aboody, R., Zhou, C., & Jara-Ettinger, J. (2021). In pursuit of knowledge: Preschoolers expect agents to weigh information gain and information cost when deciding whether to explore. *Child Development*, 92(5), 1919–1931.
- Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information?
- Alister, M., & Perfors, A. (2026). Reasoning from social information enables efficient human intelligence.
- Alister, M., Ransom, K., Connor Desai, S., Soh, E. V., Hayes, B., & Perfors, A. (2025). How convincing is a crowd? quantifying the persuasiveness of a consensus for different individuals and types of claims. *Psychological Science*, 36(7), 483–498.
- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science advances*, 6(14), eaay3539.
- Altay, S., Lyons, B. A., & Modirrousta-Galian, A. (2025). Exposure to higher rates of false news erodes media trust and fuels overconfidence. *Mass Communication and Society*, 28(2), 301–325.
- Amemiya, J., Heyman, G. D., & Gerstenberg, T. (2024). Children use disagreement to infer what happened. *Cognition*, 250, 105836.
- Anderau, G. (2023). Fake news and epistemic flooding. *Synthese*, 202(4), 106.
- Anjum, G., & Aziz, M. (2026). Responsibilized and powerless? unpacking the individualization of climate mitigation. *Energy Research & Social Science*, 133, 104585.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments.
- Avram, M., Micallef, N., Patil, S., & Menczer, F. (2020). Exposure to social engagement metrics increases vulnerability to misinformation. *arXiv preprint*.
- Baer, C., & Kidd, C. (2022). Learning with certainty in childhood. *Trends in Cognitive Sciences*, 26(10), 887–896. <https://doi.org/10.1016/j.tics.2022.07.010>

- Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4), 0064.
- 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.
- Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of cognition*, 3(1), 2.
- Baum, L. A., Danovitch, J. H., & Keil, F. C. (2008). Children's sensitivity to circular explanations. *Journal of Experimental Child Psychology*, 100(2), 146–155.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources [Place: US Publisher: American Psychological Association]. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders: Official Journal of the Movement Disorder Society*, 12(6), 1028–1034. <https://doi.org/10.1002/mds.870120629>
- Berger, J. (2011). Arousal increases social transmission of information. *Psychological science*, 22(7), 891–893.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, 81(6), 1641–1660.
- Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, 41, 15–21.
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3), 322–330.
- Bond, R. (2005). Group size and conformity. *Group processes & intergroup relations*, 8(4), 331–354.
- Braddock, K. (2022). Vaccinating against hate: Using attitudinal inoculation to confer resistance to persuasion by extremist propaganda. *Terrorism and political violence*, 34(2), 240–262.
- Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The mad model of moral contagion: The role of motivation, attention, and design in the spread of moralized content online. *Perspectives on Psychological Science*, 15(4), 978–1010.
- Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, 56(7), 1268.
- 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.

- Burnett, D. G., & Mitchell, E. (2025). Attention sanctuaries: Social practice guidelines and emergent strategies in attention activism. *Annals of the New York Academy of Sciences*, 1546(1), 5–10. <https://doi.org/10.1111/nyas.15313>
- Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child development*, 83(4), 1416–1428.
- Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children’s developing understanding that even reliable sources need to verify their claims. *Cognitive Development*, 54, 100871.
- Callara, A. L., Greco, A., Scilingo, E. P., & Bonfiglio, L. (2023). Neuronal correlates of eyeblinks are an expression of primary consciousness phenomena [Publisher: Nature Publishing Group]. *Scientific Reports*, 13(1), 12617. <https://doi.org/10.1038/s41598-023-39500-z>
- Capewell, G., Maertens, R., Remshard, M., Van Der Linden, S., Compton, J., Lewandowsky, S., & Roozenbeek, J. (2024). Misinformation interventions decay rapidly without an immediate posttest. *Journal of Applied Social Psychology*, 54(8), 441–454.
- Chan, M.-p. S., & Albarracin, D. (2023). A meta-analysis of correction effects in science-relevant misinformation. *Nature Human Behaviour*, 7(9), 1514–1525.
- Chen, A., Kumar, S. E., Varkhedi, R., & Murphy, D. H. (2024). The Effect of Playback Speed and Distractions on the Comprehension of Audio and Audio-Visual Materials. *Educational Psychology Review*, 36(3), 79. <https://doi.org/10.1007/s10648-024-09917-7>
- Christakis, D. A., Ramirez, J. S. B., & Ramirez, J. M. (2012). Overstimulation of newborn mice leads to behavioral differences and deficits in cognitive performance. *Scientific Reports*, 2, 546. <https://doi.org/10.1038/srep00546>
- Christakis, D. A., Ramirez, J. S. B., Ferguson, S. M., Ravinder, S., & Ramirez, J.-M. (2018). How early media exposure may affect cognitive function: A review of results from observations in humans and experiments in mice [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 115(40), 9851–9858. <https://doi.org/10.1073/pnas.1711548115>
- Chuey, A., & Gweon, H. (2025). Theory of minds: Early understanding of interacting minds. *Developmental Psychology*, 7(1), 91–115.
- Chung, M. (2017). Not just numbers: The role of social media metrics in online news evaluations. *Computers in Human Behavior*, 75, 949–957.
- Common Sense Media. (2019, August). New survey reveals teens get their news from social media and youtube. <https://www.common sense media.org/press-releases/new-survey-reveals-teens-get-their-news-from-social-media-and-youtube>
- Common Sense Media. (2023, May). New poll finds parents lag behind kids on ai and want rules and reliable information to help them. <https://www.common sense media.org/press-releases/new-poll-finds-parents-lag-behind-kids-on-ai>
- Compton, J., Van Der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinforma-

- tion, and conspiracy theories. *Social and Personality Psychology Compass*, 15(6), e12602.
- Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using bayesian networks. *Topics in cognitive science*, 8(1), 160–179.
- Corriveau, K. H., & Kurkul, K. E. (2014). “why does rain fall?”: Children prefer to learn from an informant who uses noncircular explanations. *Child development*, 85(5), 1827–1835.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adapted mind: Evolutionary psychology and the generation of culture*, 163, 163–228.
- Cottrell, S., Torres, E., Harris, P. L., & Ronfard, S. (2023). Older children verify adult claims because they are skeptical of those claims. *Child Development*, 94(1), 172–186.
- 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.
- C.S. Mott Children’s Hospital. (2021, October). Sharing too soon? children and social media apps. <https://mottpoll.org/reports/sharing-too-soon-children-and-social-media-apps>
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, 13(4), 148–153.
- Cutting, J. E. (2016). The evolution of pace in popular movies. *Cognitive Research: Principles and Implications*, 1(1), 30. <https://doi.org/10.1186/s41235-016-0029-0>
- Danovitch, J. H., & Alzahabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, 14(3), 499–513.
- 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.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3), 629.
- Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- 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.
- Fazio, L. K., & Sherry, C. L. (2020). The effect of repetition on truth judgments across development. *Psychological Science*, 31(9), 1150–1160.
- Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2), 117–140.

- Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction, 32*, 1–9.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99*(24), 15822–15826.
- Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental science, 11*(5), 771–777.
- Gantman, A. P., & Van Bavel, J. J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition, 132*(1), 22–29.
- Gigerenzer, G., & Brighton, H. (2009). Homo heuristics: Why biased minds make better inferences. *Topics in cognitive science, 1*(1), 107–143.
- Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 252–269. <https://doi.org/10.3758/CABN.10.2.252>
- Goldberg, M. H., van der Linden, S., Leiserowitz, A., & Maibach, E. (2020). Perceived social consensus can reduce ideological biases on climate change. *Environment and Behavior, 52*(5), 495–517.
- 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.
- Guay, B., Berinsky, A. J., Pennycook, G., & Rand, D. (2023). How to think about whether misinformation interventions work. *Nature Human Behaviour, 7*(8), 1231–1233.
- Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in cognitive sciences, 25*(10), 896–910.
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological review, 114*(3), 704.
- Handley-Miner, I. J., Pope, M., Atkins, R. K., Jones-Jang, S. M., McKaughan, D. J., Phillips, J., & Young, L. (2023). The intentions of information sources can affect what information people think qualifies as true. *Scientific Reports, 13*(1), 7718.
- 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.
- Hayes, C. (2025). *The Sirens' Call: How Attention Became the World's Most Endangered Resource*. Penguin Press.
- Heaton, B. (2024, December). 'Brain rot' named Oxford Word of the Year 2024. Retrieved January 22, 2026, from <https://corp.oup.com/news/brain-rot-named-oxford-word-of-the-year-2024/>
- Hermansen, T. K., Ronfard, S., Harris, P. L., Pons, F., & Zambrana, I. M. (2021). Young children update their trust in an informant's claim when experience tells them otherwise. *Journal of Experimental Child Psychology, 205*, 105063.

- Hermansen, T. K., Ronfard, S., Harris, P. L., & Zambrana, I. M. (2021). Preschool children rarely seek empirical data that could help them complete a task when observation and testimony conflict. *Child Development*, 92(6), 2546–2562.
- Hermansen, T. K., Viana, K. M., Harris, P. L., Engel, S., Zambrana, I. M., & Ronfard, S. (2024). Checking out the unexplained: With age, children become increasingly skeptical of surprising claims. *Developmental Psychology*, 60(10), 1761.
- Heyman, G. D., Fu, G., & Lee, K. (2007). Evaluating claims people make about themselves: The development of skepticism. *Child development*, 78(2), 367–375.
- Huang, Z., Hu, Q., & Shao, Y. (2019). Understanding others' knowledge certainty from inference and information-seeking behaviors in children. *Developmental psychology*, 55(7), 1372.
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on twitter. *Proceedings of the national academy of sciences*, 119(1), e2025334119.
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8), 589–604.
- Jarvstad, A., & Hahn, U. (2011). Source reliability and the conjunction fallacy. *Cognitive Science*, 35(4), 682–711.
- 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(10), 1541–1547.
- Kahneman, D. (1973). Attention and Effort. Retrieved January 27, 2026, from <https://philpapers.org/rec/kahaae>
- Kallioniemi, P. (2021). The role of human curation at the age of algorithms. *Journal of Digital Media & Interaction*, 4(10), 7–20.
- Kendal, R. L., Coolen, I., & Laland, K. (2009). 13 adaptive trade-offs in the use of social and personal information. *Cognitive ecology II*, 249–271.
- Kerr, J. R., & Wilson, M. S. (2018). Changes in perceived scientific consensus shift beliefs about climate change and gm food safety. *PloS one*, 13(7), e0200295.
- Kidd, C., & Birhane, A. (2023). How ai can distort human beliefs. *Science*, 380(6651), 1222–1223.
- Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, 126(1), 109–114.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5), e36399.
- Kim, J. W., Guess, A., Nyhan, B., & Reifler, J. (2021). The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, 71(6), 922–946.
- Knill, D. C., & Richards, W. (1996). *Perception as bayesian inference*. Cambridge University Press.

- Kobayashi, K. (2018). The impact of perceived scientific and social consensus on scientific beliefs. *Science Communication*, 40(1), 63–88.
- Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children’s use of true and false statements. *Psychological science*, 15(10), 694–698.
- Lang, D., Chen, G., Mirzaei, K., & Paepcke, A. (2020). Is faster better? a study of video playback speed. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 260–269. <https://doi.org/10.1145/3375462.3375466>
- 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.
- Langenhoff, A. F., Srinivasan, M., & Engelmann, J. M. (2024). Disagreement reduces overconfidence and prompts exploration in young children. *Child Development*, 95(5), 1616–1627.
- 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.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load [Publisher: Elsevier]. *Trends in Cognitive Sciences*, 9(2), 75–82. <https://doi.org/10.1016/j.tics.2004.12.004>
- Lavie, N. (2010). Attention, Distraction, and Cognitive Control Under Load [Publisher: SAGE PublicationsSage CA: Los Angeles, CA]. *Current Directions in Psychological Science*. <https://doi.org/10.1177/0963721410370295>
- Lavie, N., Hirst, A., Fockert, J., & Viding, E. (2004). Load Theory of Selective Attention and Cognitive Control. *Journal of Experimental Psychology: General*, 133, 339–354. <https://doi.org/10.1037/0096-3445.133.3.339>
- Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1822).
- Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European review of social psychology*, 32(2), 348–384.
- Lillard, A. S., & Peterson, J. (2011). The Immediate Impact of Different Types of Television on Young Children’s Executive Function. *Pediatrics*, 128(4), 644–649. <https://doi.org/10.1542/peds.2010-1919>
- Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K., & Cook, D. (2009). Using standards to improve middle school students’ accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied*, 15(4), 307.
- Lissak, G. (2018). Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research*, 164, 149–157. <https://doi.org/10.1016/j.envres.2018.01.015>
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366), 1038–1041.

- Loftus, E. F. (1979). Reactions to blatantly contradictory information. *Memory & Cognition*, 7(5), 368–374.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11), 2098.
- Lorenz-Spreen, P., Mønsted, B. M., Hövel, P., & Lehmann, S. (2019). Accelerating dynamics of collective attention [Publisher: Nature Publishing Group]. *Nature Communications*, 10(1), 1759. <https://doi.org/10.1038/s41467-019-09311-w>
- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1.
- Maheshwari, S. (2017). On youtube kids, startling videos slip past filters. *The New York Times*, 4.
- Majee, W., Anakwe, A., Onyeaka, K., & Harvey, I. S. (2023). The past is so present: Understanding covid-19 vaccine hesitancy among african american adults using qualitative data. *Journal of Racial and Ethnic Health Disparities*, 10(1), 462–474.
- March, H. A., & Simon, J. G. (1958). *Organizations*. John Wiley & Sons.
- Marti, L., Conover, A., & Kidd, C. (2021). “fringe” beliefs aren’t fringe.
- Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children’s vigilance towards deception. *Cognition*, 112(3), 367–380.
- McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. (2024). Misinformation exploits outrage to spread online. *Science*, 386(6725), 991–996.
- Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.
- Mercier, H., & Miton, H. (2019). Utilizing simple cues to informational dependency. *Evolution and Human Behavior*, 40(3), 301–314.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Harvard university press.
- Mills, C. M., & Elashi, F. B. (2014). Children’s skepticism: Developmental and individual differences in children’s ability to detect and explain distorted claims. *Journal of experimental child psychology*, 124, 1–17.
- Mills, C. M., Sands, K. R., Rowles, S. P., & Campbell, I. L. (2019). “i want to know more!”: Children are sensitive to explanation quality when exploring new information. *Cognitive Science*, 43(1), e12706.
- Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General*, 152(9), 2411.
- Nazaryan, A. (2026). They Want to Influence You to Do ... Nothing. *The New York Times*. Retrieved February 2, 2026, from <https://www.nytimes.com/2026/01/01/style/boredom-online-trend-influencers-tiktok.html>

- Nguyen, L., Walters, J., Paul, S., Monreal Ijurco, S., Rainey, G. E., Parekh, N., Blair, G., & Darrah, M. (2025). Feeds, feelings, and focus: A systematic review and meta-analysis examining the cognitive and mental health correlates of short-form video use. *Psychological Bulletin*, 151(9), 1125–1146. <https://doi.org/10.1037/bul0000498>
- O'Donnell, R., & Chan, J. C. K. (2025). Does blatantly contradictory information reduce the misinformation effect? a registered report replication of Loftus (1979). *Legal and Criminological Psychology*, 30(1), 3–26. <https://doi.org/https://doi.org/10.1111/lcrp.12242>
- Oktar, K., Sumers, T., & Griffiths, T. (2024). A rational model of vigilance in motivated communication. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46.
- Orchinik, R., Martel, C., Rand, D. G., & Bhui, R. (2025, March). Adaptive intuitions shape susceptibility to misinformation. https://doi.org/10.31234/osf.io/q7r58_v2
- Orticio, E., Gao, S., & Kidd, C. (2025). Children learn the meaning of ambiguous evidence from third-party belief revision. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 47.
- Orticio, E., & Kidd, C. (2024). Shifting your opinion makes you change your factual beliefs without evidence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46.
- Orticio, E., Martí, L., & Kidd, C. (2022). Social prevalence is rationally integrated in belief updating. *Open Mind*, 6, 77–87.
- Orticio, E., Meyer, M., & Kidd, C. (2024). Exposure to detectable inaccuracies makes children more diligent fact-checkers of novel claims. *Nature Human Behaviour*, 8(12), 2322–2329.
- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental psychology*, 43(5), 1216.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50.
- Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality*, 88(2), 185–200.
- Perfors, A., & Navarro, D. J. (2019). Why do echo chambers form? the role of trust, population heterogeneity, and objective truth. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 41.
- Perfors, A., Navarro, D. J., & Shafto, P. (2018). Stronger evidence isn't always better: A role for social inference in evidence selection and interpretation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 40.
- Plate, R. C., Shutts, K., Cochrane, A., Green, C. S., & Pollak, S. D. (2021). Testimony bias lingers across development under uncertainty. *Developmental psychology*, 57(12), 2150.

- Rahnev, D. (2025). A comprehensive assessment of current methods for measuring metacognition. *Nature Communications*, 16(1), 701. <https://doi.org/10.1038/s41467-025-56117-0>
- Ransom, K. J., Perfors, A., & Stephens, R. (2021). Social meta-inference and the evidentiary value of consensus. *Proceedings of the annual meeting of the cognitive science society*, 43(43).
- Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the national academy of sciences*, 118(26), e2024292118.
- Reese, D., Pawluk, D., & Taylor, C. (2016, December). Engaging Learners Through Rational Design of Multisensory Effects. In *Emotions, Technology, and Design* (pp. 103–127). <https://doi.org/10.1016/B978-0-12-801872-9.00006-5>
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T., & Laland, K. N. (2010). Why copy others? insights from the social learning strategies tournament. *Science*, 328(5975), 208–213.
- Richardson, E., & Keil, F. (2022). “he only changed his answer because they shouted at him”: Children use affective cues to distinguish between genuine and forced consensus. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44).
- Rippon, S. (2024). Evidential incognizance. *Acta Analytica*, 39(4), 663–676.
- Rodriguez, A. (2022, July). Youtube kids is giving parents more control over what their kids watch. <https://qz.com/1262977/youtube-kids-is-launching-a-mode-curated-by-humans-not-just-algorithms>
- Rouault, M., & Fleming, S. M. (2020). Formation of global self-beliefs in the human brain. *Proceedings of the National Academy of Sciences*, 117(44), 27268–27276. <https://doi.org/10.1073/pnas.2003094117>
- Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science*, 26(5), 521–562.
- Ruggeri, A. (2022). An introduction to ecological active learning. *Current Directions in Psychological Science*, 31(6), 471–479.
- Ryan, S. (2018). Epistemic environmentalism. *Journal of Philosophical Research*.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *science*, 274(5294), 1926–1928.
- Salles, A., Ais, J., Semelman, M., Sigman, M., & Calero, C. I. (2016). The metacognitive abilities of children and adults. *Cognitive Development*, 40, 101–110.
- Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity spreads more than positivity on twitter after both positive and negative political situations. *Affective Science*, 2(4), 379–390.
- Shepherd, S. S., Orticio, E., Castañeda Barba, M., & Kidd, C. (2026). The brain rot effect: The case for adaptive attentional bandwidth. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 48.
- Shepherd, S. S., & Kidd, C. (2024). Visual engagement is not synonymous with learning in young children. *Proceedings of the Annual Meeting of the Cognitive Science So-*

- ciety*, 46(0). Retrieved September 6, 2024, from <https://escholarship.org/uc/item/0wz74769>
- Simon, H. A., et al. (1990). Invariants of human behavior. *Annual review of psychology*, 41(1), 1–20.
- Simon, H. A., Deutsch, K. W., & Shubik, M. (1971). Designing organizations for an information-rich world. *Computers, communications, and the public interest*.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological bulletin*, 105(1), 131.
- Smith, A. C., Woerner, J., Perera, R., Haeny, A. M., & Cox, J. M. (2022). An investigation of associations between race, ethnicity, and past experiences of discrimination with medical mistrust and covid-19 protective strategies. *Journal of racial and ethnic health disparities*, 9(4), 1430–1442.
- Smith, D. G. (2023). How to Focus Like It's 1990. *The New York Times*. Retrieved February 2, 2026, from <https://www.nytimes.com/2023/01/09/well/mind/concentration-focus-distraction.html>
- Sommer, J. (2026). In the mind or in the world? types of beliefs and the locality of evidence. *Psychological Inquiry*, 37(1), 1–26. <https://doi.org/10.1080/1047840X.2026.2650209>
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & language*, 25(4), 359–393.
- Stephens, E. C., & Koenig, M. A. (2015). Varieties of testimony: Children's selective learning in semantic versus episodic domains. *Cognition*, 137, 182–188.
- Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The Endogenous Eyeblink [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.1984.tb02312.x>]. *Psychophysiology*, 21(1), 22–33. <https://doi.org/10.1111/j.1469-8986.1984.tb02312.x>
- Subedar, A., & Yates, W. (2017). The disturbing youtube videos that are tricking children. *BBC News*, 27.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Szegöfi, A., Altay, S., & Heintz, C. (2026). “flooding” using disinformation. *Available at SSRN 6216050*.
- Szegöfi, Á., & Heintz, C. (2022). Institutions of epistemic vigilance: The case of the newspaper press. *Social Epistemology*, 36(5), 613–628.
- Tharumalingam, T., Roberts, B. R. T., Fawcett, J. M., & Risko, E. F. (2025). Increasing Video Lecture Playback Speed Can Impair Test Performance – a Meta-Analysis. *Educational Psychology Review*, 37(2), 35. <https://doi.org/10.1007/s10648-025-10003-9>
- The Friends of Attention. (2026, January). *Attensity!: A Manifesto of the Attention Liberation Movement*. Crown. <https://www.penguinrandomhouse.com/books/782387/attensity-by-the-friends-of-attention/>
- Thielmann, I., & Hilbig, B. E. (2023). Generalized dispositional distrust as the common core of populism and conspiracy mentality. *Political Psychology*, 44(4), 789–805.

- Todd, P. M., & Gigerenzer, G. (2012). *Ecological rationality: Intelligence in the world*. OUP USA.
- Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B: Biological Sciences*, 281(1776).
- Toelch, U., & Dolan, R. J. (2015). Informational and normative influences in conformity from a neurocomputational perspective. *Trends in cognitive sciences*, 19(10), 579–589.
- Tong, Y., Wang, F., & Danovitch, J. (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, 23(2), e12895.
- Tong, Y., Wang, F., Danovitch, J., & Wang, W. (2022). When the internet is wrong: Children's trust in an inaccurate internet or human source. *British Journal of Developmental Psychology*, 40(2), 320–333.
- Udry, J., & Barber, S. J. (2024). The illusory truth effect: A review of how repetition increases belief in misinformation. *Current Opinion in Psychology*, 56, 101736.
- Upton-Clark, E. (2024, November). 'PDF to brainrot' tools are bringing TikTok visuals to Gen Alpha's academics. Retrieved February 1, 2026, from <https://www.fastcompany.com/91235324/pdf-to-brainrot-tools-are-bringing-tiktok-visuals-to-gen-alphas-academics>
- Van Der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature medicine*, 28(3), 460–467.
- Van der Linden, S., Leiserowitz, A., & Maibach, E. (2019). The gateway belief model: A large-scale replication. *Journal of Environmental Psychology*, 62, 49–58.
- Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global challenges*, 1(2), 1600008.
- Van Prooijen, J.-W., Krouwel, A. P., & Pollet, T. V. (2015). Political extremism predicts belief in conspiracy theories. *Social psychological and personality science*, 6(5), 570–578.
- Vanderbilt, K. E., Liu, D., & Heyman, G. D. (2011). The development of distrust. *Child development*, 82(5), 1372–1380.
- Vanderbilt, K. E., Ochoa, K. D., & Heilbrun, J. (2018). Consider the source: Children link the accuracy of text-based sources to the accuracy of the author. *British Journal of Developmental Psychology*, 36(4), 634–651.
- Vlasceanu, M., Goebel, J., & Coman, A. (2020). The emotion-induced belief amplification effect. *Proceedings of the annual meeting of the cognitive science society*, 42.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146–1151.
- Wang, J., Yang, Y., Macias, C., & Bonawitz, E. (2021). Children with more uncertainty in their intuitive theories seek domain-relevant information. *Psychological Science*, 32(7), 1147–1156.
- Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity [Publisher: The

- University of Chicago Press]. *Journal of the Association for Consumer Research*, 2(2), 140–154. <https://doi.org/10.1086/691462>
- Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological science*, 29(7), 1159–1177.
- Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive science*, 42(1), 168–187.
- Williams, D. (2023). The fake news about fake news. *Boston Review*, 7.
- Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Knopf Doubleday Publishing Group. Retrieved January 24, 2024, from <http://ebookcentral.proquest.com/lib/berkeley-ebooks/detail.action?docID=6040492>
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, 105(13), 5012–5015.
- Xu, F., & Tenenbaum, J. B. (2007). Sensitivity to sampling in bayesian word learning. *Developmental science*, 10(3), 288–297.
- Xu, S., Shtulman, A., & Young, A. G. (2022). Can children detect fake news? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44(44).
- Yousef, A. M. F., Alshamy, A., Tlili, A., & Metwally, A. H. S. (2025). Demystifying the New Dilemma of Brain Rot in the Digital Era: A Review. *Brain Sciences*, 15(3). <https://doi.org/10.3390/brainsci15030283>
- Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science*, 30(8), 1195–1204.